T889

Problem solving and improvement: quality and other approaches

The Open University

Study guide

Block 1 Introduction

Block 2 Statistics

T889

Problem solving and improvement: quality and other approaches

The Open University

Study guide

Block 1 Introduction

Block 2 Statistics

This publication forms part of an Open University course T889 *Problem solving and improvement: quality and other approaches*. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom: tel. +44 (0)845 300 60 90, email general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at http://www.open.ac.uk where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials visit http://www.ouw.co.uk, or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure. tel. +44 (0)1908 858793; fax +44 (0)1908 858787; email ouw-customer-services@open.ac.uk

FSC www.fsc.org MIX Paper from responsible sources FSC® C016278

The Open University
Walton Hall, Milton Keynes
MK7 6AA

First published 2007.

Edited and designed by The Open University.

Typeset by SR Nova Pvt. Ltd, Bangalore, India.

Printed and bound in the United Kingdom by Halstan Printing Group, Amersham.

ISBN 978 0 7492 2349 6

1.1

COURSE TEAM LIST

Joyce Fortune, course team chairman and main author

Deborah Mairs, course manager

Caroline McLernon, course secretary

Critical readers

Malcolm Fowles

John Hughes

John Swanwick

Contributing authors

Brian Bloomfield

Malcolm Fowles

John Hughes

Nick Oliver

Media production staff

Jill Alger

Annette Booz

Sophia Braybrooke

Philippa Broadbent

Jane Bromley

Kim Dulson

Vicky Eves

Carol Houghton

Sian Lewis

Katie Meade

Lara Mynors

Deana Plummer

DVD Programme production

The credits for Programmes 1, 2 and 5 are shown at the end of the programmes.

Programmes 3 and 4 were made by Twofour Communications

Academic Consultant: Joyce Fortune

Producer/Director: Hannah Craven

Assistant Producer: Lucinda Jellier

Camera: Steve Horrocks and Jamie Ryan

Sound: Malcolm Flynn and Curig Huws

Offline Editor: Stuart Robinson

Online Editor: Sam Storey

External assessor

Prof. Nigel Slack, Warwick Business School, The University of Warwick

Study guide

CONTENTS

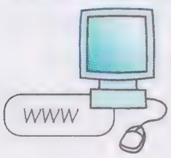
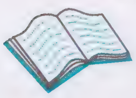
1 WELCOME TO THE COURSE

Welcome to T889 *Problem solving and improvement: quality and other approaches.* On behalf of the course team may I wish you well with your studies. We hope you find the course stimulating and productive and that it proves to be of long-term benefit.

2 COURSE COMPONENTS

The main spine of the course is made up of five blocks of study. As you work through them you will be directed to other resources. When this happens you will see an icon in the margin to indicate the type of resource you will be using. These icons are shown in Table S.1.

Table S.1 The margin icons

Icon	Resource
	Course website
	Offprints
	Statistical software/*Computer Exercises Booklet*
	The programmes on your DVD

The Study Calendar on the course website uses the blocks as the main milestones for managing your rate of study in the first two-thirds of the course. During the final third you undertake a project and write a project report that acts as your end-of-course assessment (ECA).

T889 is a 30 CATS points course (CATS stands for the UK national Credit Accumulation and Transfer Scheme). Broadly speaking, the scheme equates 1 'credit point' with 10 hours of learning effort or notional learning time, so the course as a whole should require you to commit approximately 300 hours of work. To put it another way, you need to spread almost eight weeks of full-time study over a period of a little more than five months. The Study Calendar assumes you will be studying at a steady pace throughout the course but of course you are free to vary your rate of study to suit yourself.

However, there are four fixed points on the calendar – the cut-off dates for each of the three tutor-marked assignments (TMAs) and the date for submitting your ECA – and it is essential that you accommodate these.

The course website

You have been sent information on how to access the course website. If you haven't already done so, I suggest you access it as soon as possible. The Study Calendar is located there, and the course website provides you with a wide variety of other online resources and documents in electronic form, including a guide on how to use the website itself. It is also your route into the course forum and the place where you will find the information you need for your project and your ECA. Any news items will also be posted on the course website.

The website also provides you with access to the Open University library and a selection of resources that we think might be of particular interest, including links to electronic journals you might like to browse.

There is no need to start printing a lot of the documents available on the course website. You may find it useful to print some of them if you start using them to a significant extent, but you shouldn't find yourself printing reams.

The blocks of study

The main blocks are:

Block 1 *Introduction*
Block 2 *Statistics*
Block 3 *Techniques*
Block 4 *Methods and approaches*
Block 5 *Managing problem solving and improvement.*

After Block 5 you will find an *Index of techniques*, which will act as an aide-memoire both in your assignments for this course and in your future studies or career.

As you study the blocks, the main form of interaction with them is through the activities and, in the case of Blocks 2 and 3, the computer exercises. Some of the activities are designed to allow you to check your understanding of the material and others are there to help you to develop skills. Many of the activities have 'answers' at the end of the block. Do be aware, however, that the inverted commas around the word answers are important. Sometimes there are right and wrong answers, especially in Block 2, but often there is a large degree of subjectivity in the responses you are making. For example, suppose I ask you to draw a systems map (don't worry if you haven't met systems maps before – you will in Block 3). In judging your attempt, there is a right or wrong aspect in relation to how closely you have followed the systems maps conventions and whether you have supplied a title spelling out the type of diagram and the name of the system depicted, but when it comes

to judging the content there is room for interpretation. Thus, looking at Figure S.1, part (a) is wrong because it breaches the convention that components are placed in blobs, but whether (b) or (c) is a truer representation depends on whether you perceive component 'ghi' to be part of the system or part of the environment. Where the degree of subjectivity is very large or the question is based entirely on your own experiences, there may be no 'answer' in the back at all. In these cases it is very useful to engage with your tutor and your fellow students in the course online forum, and there will be a website icon in the margin to remind you to do this.

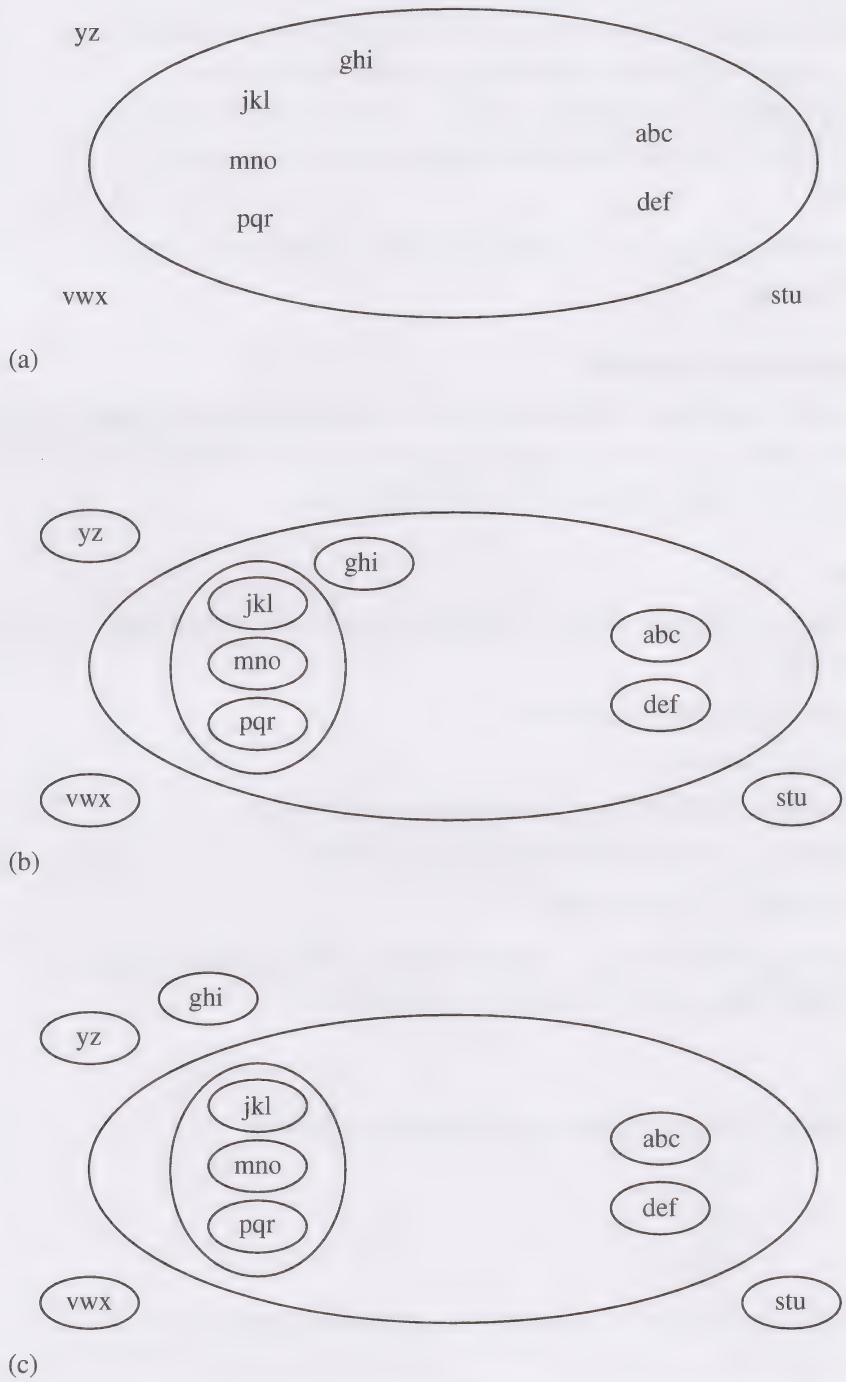


Figure S.1 Three responses to a request to draw a systems map

Offprints

Some journal articles that are extremely relevant to the course content have been selected as offprints. The places where you are recommended to read them are indicated with the offprint icon in the margin. They are provided electronically rather than in print so that we can ensure the selection remains relevant and make any necessary changes as close as possible to the start of the course.

Statistical software

A CD containing a leading-edge statistical analysis software package is supplied with this course. The instructions for loading the software are on the CD. If you have problems installing it please contact the OU Computing Helpdesk. Information on the various ways to do this can be found at http://www.open.ac.uk/students/helpdesk/, via the 'Contact Us' link.

The CD also contains files you will need in order to undertake some of the computer exercises.

Computer Exercises Booklet

For ease of use, the computer exercises have been printed in a separate booklet. You will need to do these exercises during your study of Blocks 2 and 3 at the places indicated in the block texts.

Programmes

The five programmes that you will be watching during the course are supplied on a DVD.

1 Managing processes: SPC in action
2 Recognising excellence
3 Problem solving in action: Six Sigma at ScottishPower
4 Problem solving in action: transformation at COSi
5 Customers, quality & competition

You will be watching Programme 1 during Block 3, Programmes 2 and 3 during Block 4 and Programmes 4 and 5 during Block 5.

Assessment

The course has three TMAs. These are weighted as follows:

- TMA 01 30%
- TMA 02 40%
- TMA 03 30%

Together the TMAs make up the continuous assessment component of the assessment strategy and account for 50% of your course result. The ECA

(a project report with a maximum length of 3000 words) accounts for the other 50%.

The assignments and information about the ECA are available on the course website. I suggest you take a little time now to browse through these, noting the cut-off dates for the three TMAs, and looking at the sorts of assessment task you will be asked to perform. The assignment section also contains other useful information such as the referencing system you are expected to use in your work and how to avoid plagiarism.

3 COURSE CONTENT

The final report of the Leitch review of skills makes an important point:

> High skills are becoming increasingly important in the global economy. They drive growth, facilitate innovation and are crucial for world-class management and leadership.
>
> (Leitch, 2006, p. 66)

Alongside literacy and numeracy, one of the generic skills that is particularly relevant in this context is problem solving. However, it is clear that problem-solving skills are not as well developed or widely deployed as they might be. Indeed, one of the key findings of the National Employers Skills Survey for 2005 (Learning and Skills Council, 2006) was that problem-solving skills were lacking in applicants in a third of skills-shortage vacancies.

This course aims to help you in two ways. The first is to enhance your own skills in problem solving and improvement, and the second is to help you develop the skills of other people and to manage their problem-solving and improvement activities.

One point I should like to emphasise is that this course is not seeking to push you down one particular route to solving problems and achieving improvements. Over time, improvement in particular has been dogged by the 'flavour of the month' or 'passing fad' accusation. In one sense, taking up initiatives that fall by the wayside after early gains (picking just the low-hanging fruit, if you like) is still beneficial, provided that the costs do not exceed the investments, but there is a big potential danger in this approach. It is the danger of initiative fatigue, which leads to weariness, cynicism and diminishing returns. It is therefore essential to survey the different techniques, methods and approaches that are available, look at the needs of your organisation and map out your own way forward. This course helps you to do this. For example, one of the issues discussed in Block 1 is why it is important to take situational factors into account when looking at problems and opportunities.

As you study the material you will see that most of the techniques you will be studying are not new – even if they are sometimes made to appear new! As Box S.1 shows, Taylor and Ford laid down many of the principles that underpin problem solving and improvement a century ago. As you study the course, look back at these principles and see how they are being applied in the methods and approaches that you meet. Then ask yourself: What separates this method or approach from the rest? What makes it special? It is worth abandoning the tried and tested generic methods covered in Block 1 only if the something special offered by another method or approach is particularly applicable to your organisation and the problems and opportunities it faces.

BOX S.1 THE EVOLUTION OF TOP-DOWN AND BOTTOM-UP CONCEPTS OF PROBLEM SOLVING AND IMPROVEMENT AS BUSINESS THEMES

Entrepreneurs and industrialists have always looked for ways of reducing costs and, as a consequence, of increasing profits but the idea of an explicit improvement agenda may be traced to the later decades of the nineteenth century. One of the pioneers of the American steel industry, Andrew Carnegie, writing of the period 1863–68 stated: 'The surest foundation of a manufacturing concern is quality. After that, and a long way after, comes cost' (Carnegie, 1920). But perhaps the greatest influence on problem solving and improvement during the period towards the end of the nineteenth and beginning of the twentieth centuries, was exerted by a loose group of industrialists and individuals who were members of the American Society of Engineers. (Nowadays these people would be called management consultants although the term was not used at the time.) These people discussed and made contributions to the resolution of issues that continue to preoccupy managers today. Pre-eminent among the group was F.W. Taylor. He developed a set of principles to which he gave the name 'scientific management' and which were to become what might be termed the dominant paradigm for productivity improvement in the twentieth century (Taylor, [1911] 1998).

Taylor's aims were: to raise awareness of the loss that was being suffered to 'the whole country' through inefficiencies in working practice; to convince managers of the need for a systematic approach to remedy inefficiency; and to prove that management is a science which should have a foundation of laws, rules and principles. He summarised scientific management as:

- science, not rule of thumb
- harmony, not discord
- cooperation, not individualism
- maximum output, in place of restricted output.

(Taylor, [1911] 1998, p. 74)

Taylor's ideas fundamentally altered the way in which work responsibilities were split between management and workers. The traditional approach was for management to decide *what* needed to be done, leaving the worker to work out *how* best to do the work. In between these two positions was a disputed territory of *how much* was to be done and *when* it was to be done by, in other words throughput. To gain control of this industrial no man's land Taylor proposed that determining the best work methods should become the responsibility of management. He proposed new duties of management which would be grouped under four principles:

> *First.* They [management] develop a science for each element of a man's [*sic*] work, which replaces the old rule-of-thumb method.
>
> *Second.* They scientifically select and then train, teach, and develop the workman, whereas in the past he chose his own work and trained himself as best he could.
>
> *Third.* They heartily cooperate with the men so as to assure all of the work being done in accordance with the principles of the science which has been developed.
>
> *Fourth.* There is almost an equal division of the work and the responsibility between the management and the workmen. The management take over all work for which they are better fitted than the workmen, while in the past almost all of the work and the greater part of the responsibility were thrown upon the men.
>
> (Taylor, [1911] 1998, pp. 15–16)

This new split of duties in which management, under the first principle, assumed the responsibility for deciding on and implementing methods of work, has persisted until the present day. It gave rise to time and motion study, industrial or production engineering, organisation and methods, Six Sigma and other approaches, all of which are or can be top-down in their application, in that they are sanctioned by senior managers who usually delegate their implementation to a group of experts.

At the same time that Taylor and his co-workers were developing their ideas of top-down improvement, another famous American industrialist, Henry Ford, was adopting a more inclusive policy. Ford's approach was more pragmatic. He wanted to mobilise the whole workforce in the cause of continuous improvement:

> Everyone in the place reserves an open mind as to the way in which every job is being done. If there is any fixed theory – any fixed rule – it is that no job is being done well enough. The whole factory management is always open to suggestion,

> and we have an informal suggestion system by which any workman [*sic*] can communicate any idea that comes to him and get action on it.
>
> (Ford and Crowther, [1922] 2003, p. 100)

This quotation indicates the fundamental difference between Taylor's ideas of scientific management and Ford's 'open mind' way of regarding improvement. For Taylor:

- there is a best way to do a job, which
- can be established through scientific experiments, which
- are undertaken by experts under the direction of management.

Thus scientific management, for all its avowed progressive intent, has a static character, for how is it possible to improve on a method that has been established as being the *one best way* to do the job?

In contrast, Henry Ford's philosophy may be summarised as:

- the way in which a job is being done can always be improved, and
- everyone in the organisation can make suggestions for improvement.

In the decades between the founding of the Ford factories and the start of World War II, this bottom-up or inclusive way of achieving performance improvement was largely forgotten, not least by the Ford company itself, and Taylor's ideas came to dominate the way in which improvement was achieved.

4 LEARNING OUTCOMES

The learning outcomes of the course that will be assessed in the TMAs and ECA are as follows.

Knowledge and understanding

Demonstrate knowledge and understanding of:

- Concepts associated with the nature of problems, the sources of solutions, breakthrough and continuous improvement.
- Statistical techniques relevant to problem solving and improvement.
- Concepts, approaches, methods and techniques in relation to:
  - identification of improvement opportunities
  - selection of problems on which to operate
  - investigation and description of problem situations using quantitative and qualitative methods
  - investigation of symptoms and causes
  - generation of solutions and selection between them
  - generation of implementation plans that take account of contexts and constraints.

Cognitive skills

- Investigate, analyse, think critically, evaluate and synthesise information relating to problems and opportunities for improvement from a range of appropriate sources.

Key skills

- Communicate effectively using written and graphical presentations as appropriate, producing detailed analyses of problems and opportunities for improvement.
- Draw lessons from investigations and analyses of problems and opportunities for improvement.
- Work independently, reflecting on your own actions and thoughts, and making effective use of constructive feedback.

Practical and/or professional skills

- Select the most appropriate methods for problem solving and/or improvement in a familiar situation.
- Participate in the application of a wide variety of investigative and problem-solving/improvement techniques.

REFERENCES

Carnegie, A. (1920) *Autobiography*, Ch 9, http://www.wordowner.com/carnegie/chapter9.htm (accessed December 2006).

Ford, H. and Crowther, S. ([1922] 2003) *My Life and Work*, Whitefish, MT, Kessinger Publishing.

Leitch, S. (2006) *Prosperity for All in the Global Economy – World Class Skills*, London, The Stationery Office.

Learning and Skills Council (2006) *National Employers Skills Survey 2005: Key Findings*, Coventry, Learning and Skills Council, http://www.lsc.gov.uk (accessed 13 April 2007).

Taylor, F. W. ([1911] 1998) *The Principles of Scientific Management*, New York, Dover Publications.

Block 1 Introduction

CONTENTS

AIMS

The aims of Block 1 are to:

- set the scene for the course by looking at what is meant by the terms 'problem', 'opportunity', 'improvement' and 'solution'
- identify different classes of problem and categories of improvement
- look at the characteristics of different types of approach to problem solving and improvement
- introduce a range of generic approaches to problem solving and improvement.

LEARNING OUTCOMES

After studying Block 1 you should be able to:

- distinguish between different classes of problem
- distinguish between different categories of improvement
- distinguish between different types of approach to problem solving and improvement
- demonstrate familiarity with three generic methods for problem solving and improvement.

1 INTRODUCTION

Until the 1980s, problem solving and improvement in commerce and industry, by and large, fell into two types: firefighting and planned. The former was carried out by managers, usually reacting to crises, and the latter was undertaken by consultants or internal specialists trained in method study, work study, organisation and methods (O&M), operations research (OR) and the like. 'Quick fixes' were common, especially where firefighting was concerned; analysis and the use of techniques were confined to the specialists; and follow-up work to check that problems really had been solved and that the improvements promised were being delivered was virtually non-existent throughout.

In the 1980s things started to change. The trigger for this change was the explosion of interest in modern quality management. One of its earliest manifestations was the widespread introduction of an early example of 'learn from Japan': quality circles. The quality circle movement had started in Japan, was boosted in 1962 with the publication of the first issue of *Gemba-To-QC* (*Quality Control for the Foremen*), and was subsequently credited with much of the improvement that transformed the country's manufacturing industry in the 1960s and 1970s. The circles were small groups of people, usually between five and ten in number, meeting voluntarily to try to improve quality and productivity in their work areas. For a time quality circles became very popular in the West too, and there were widespread reports of their successes. Before long, however, their ability to deliver in the longer term was being questioned and many of those that had been set up fell by the wayside. The quality circle phenomenon then gave way to the more closely managed and formally organised concept of the quality improvement team and to various other arrangements I shall be discussing later in the course. Nevertheless, quality circles remain significant in the problem-solving and improvement story because they represent the first attempt to expand problem-solving and improvement activities beyond managers and specialists.

Quality circles are also interesting in another respect: they quickly divided into two sorts that typify the extremes of the five types of problem solving shown in Box 1.1. In some organisations quality circle members were trained from the start to use a variety of techniques and to work through a systematic, structured approach, often with formal minutes of meetings and reports of their investigations and findings. Other organisations offered no training but relied on a mix of experience and intuition among circle members to generate recommendations. The solutions the latter type suggested were often spectacularly successful but equally often they failed to deliver material benefits to the organisation, and these untrained circles ran out of steam quickly.

BOX 1.1 FIVE APPROACHES TO PROBLEM SOLVING

[Problem solving] can be approached:

1. purely intuitively without careful reflection about the problem
2. through routine recourse to procedures used in the past
3. by adopting unquestioningly the solutions suggested by experts
4. by choosing at random
5. on the basis of systematic rational thought supported by relevant information.

(Source: Grünig and Kühn, 2005, pp. 7–8)

The rationale for this course rests on two foundations. First, significant advantages to organisations result from adopting a structured approach to problem solving and improvement. Second, in order to benefit fully from these advantages organisations must have the skill set needed to construct and maintain an internal infrastructure that will ensure the approach is used effectively. The main advantages of adopting a structured approach are that it:

- gives an identity to the problem-solving or improvement exercise, thus making it more likely that it will be followed through to a conclusion
- legitimises the use of time and other resources on this type of activity
- imposes rigour that ensures analysis is carried out, potential solutions are generated and a selection process is undertaken that leads to recommendations
- makes better use of the available knowledge base. In some cases this knowledge may be available internally; for others it may be necessary to seek it outside the organisation
- facilitates effective group working
- is more likely to reveal unexpected connections, contradictions and invalid assumptions
- delivers transparency in decision making
- makes it easier to implement the actions decided upon
- gives the outcomes of the exercise greater validity and therefore makes it easier to defend them
- provides a way of structuring training in problem solving.

In Blocks 2, 3 and 4 you will find a vast array of techniques that can be used as part of a structured approach. You will also find a range of methods from the very simple to the highly formal and stylised. It is not the purpose of the course to champion one method over another; instead it aims to put

forward a number of options and provide you with a means of selecting the most appropriate for a particular set of problems or opportunities, or for a specific organisational setting. This block ends by introducing some generic methods but before then I shall set the scene by looking at the different types of problems and improvement.

2 PROBLEMS AND IMPROVEMENT

In this course I shall often use the words 'problem' and 'improvement' but you will see the word 'opportunity' much less frequently. However, if I were to be pedantic, 'opportunity' would appear much more often. The reason for this is that there are two types of opportunity: those that involve the creation of something new; and those that signal a chance to improve. Almost every method and technique that can be used to try to solve a problem can also be used to try to create the conditions needed to benefit from an opportunity. Equally, every attempt to bring about improvement must be underpinned by the belief that an opportunity for improvement exists. Therefore, although I will not be using the word often, please take it as read that when I refer to 'problem' I usually mean 'problem/opportunity' and when I talk about improvement I am assuming that an opportunity for improvement is believed to exist.

Perhaps the main distinguishing features that separate problem solving from improvement are the triggers that cause these types of activity to be undertaken and the urgency with which solutions need to be found. Problem solving is usually triggered by the perception of a sudden or gradual deterioration in performance from the expected level, whereas an improvement exercise follows a desire to increase performance above the current, expected level. Where urgency is concerned, the distinction is not always clear-cut. For example, an opportunity to expand the range of services offered may require improvements to the existing situation as well as the introduction of new facilities, and in those circumstances the urgency with which improvements must be sought may be very great. Improvement can also be urgent when it is needed to remain competitive.

As you will see later in the course, there are usually very few differences between the methods and techniques that can be used for problem solving and those that can be used for improvement. However, greater differences in their applicability can be found when different types and levels of problem or opportunity for improvement are considered. For that reason I shall look at a number of ways of classifying problems, improvement and solutions.

2.1 The nature of problems

In everyday conversation we would expect the words 'a simple problem' to mean one that is easy to solve, but in much of the literature of problem solving, instead of denoting the opposite of difficult, simple means the opposite of complex. Flood and Jackson (1991), for example, in their book *Creative Problem Solving*, draw distinctions similar to those shown in Table 1.1.

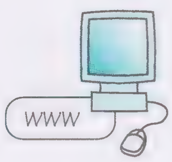
Table 1.1 Categories of problems

Simple problem	**Complex problem**
Small number of elements	Large number of elements
Few interactions between elements	Many interactions between elements
Attributes of elements are predetermined	Attributes of elements are not predetermined
Interaction between elements is highly organised	Interaction between elements is loosely organised
Well-defined laws govern behaviour	Ill-defined laws govern behaviour
No evolution over time	Evolves over time
Single set of goals	Complex set of goals
Strong boundary	Weak boundary

(Source: adapted from Flood and Jackson, 1991, pp. 33–4)

As you can see from the table, Flood and Jackson are not only concerned with the problem itself. By referring to goals and a boundary they are starting to consider the situation within which the problem exists. This notion of a problem situation is very important because it is one of the main reasons why standard 'solutions' often don't work. 'Not invented here' and 'reinventing the wheel' are pejorative terms which imply that refusal to implement solutions from elsewhere is never justified, but it is sometimes the case that only 'home-grown' solutions will be successful. Although one problem can appear to be the same as another, its situation may mean that a very different solution is required.

ACTIVITY 1.1



From your own work situation or a situation you know well, choose two problems – one that was/would be simple to solve and one that was/would be difficult to solve. Now assess each problem against the characteristics in Table 1.1. Where do your problems fit in Figure 1.1 (overleaf)? Is your simple-to-solve problem simple and simple or simple and complex? Is your difficult-to-solve problem simple or complex? Are there oddities in the categorisation such as one dimension at odds with the rest?

Now think about the extent to which standard or home-grown solutions were (or could be) applied to each problem, and estimate their degree of success. In your cases are standardisation and success influenced more by the problem's simplicity/complexity or by the problem's context?

One of the main causes of differences between contexts is the people involved and the political/cultural climate in which they operate. Looking at the relationships within groups of people, Flood and Jackson identify three

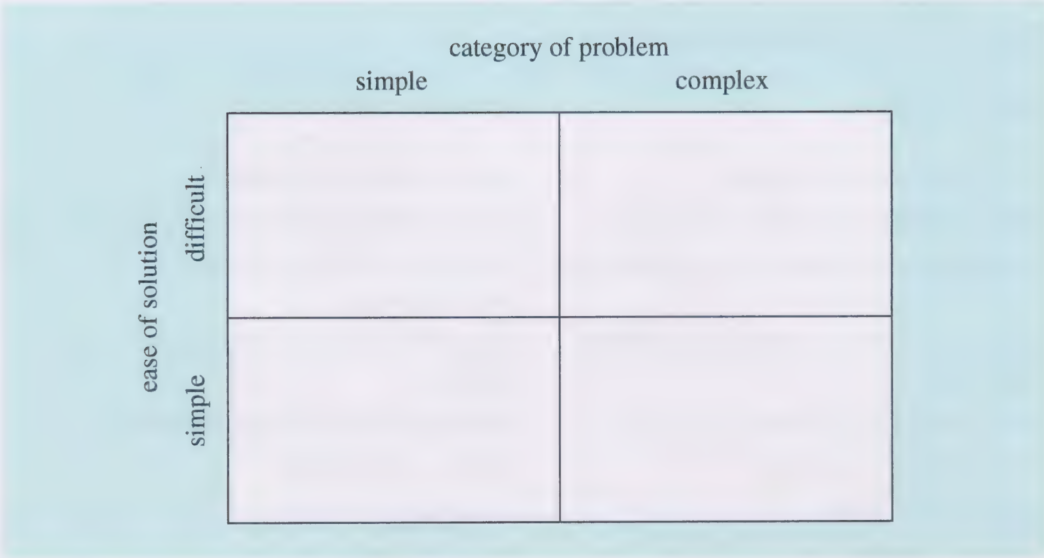


Figure 1.1 A two-way categorisation

types of situation: unitary, pluralist and coercive. The characteristics of these are as follows (Flood and Jackson, 1991, pp. 34–5):

Unitary

where the people:

- share common interests
- have values and beliefs that are highly compatible
- largely agree on ends and means
- all participate in decision making
- act in accordance with agreed objectives.

Pluralist

where the people:

- have a basic compatibility of interest
- have values and beliefs that diverge to some extent
- do not necessarily agree on ends and means, but compromise is possible
- all participate in decision making
- act in accordance with agreed objectives.

Coercive

where the people:

- do not share common interests
- have values and beliefs that are likely to conflict
- do not agree on ends and means and 'genuine' compromise is not possible
- are divided and some coerce others to accept decisions
- are unable to agree over objectives (in the present circumstances).

Box 1.2 shows extracts from four authors who take Flood and Jackson's distinction between simple and complex further. Although the four use different terminology they are essentially agreeing with the point I made earlier: different categories of problem call for different forms of problem-solving activity.

BOX 1.2 DICHOTOMIES OF PROBLEMS

Ackoff ... messes versus problems

According to Ackoff (1979) 'Managers are not confronted with problems that are independent of each other, but with dynamic situations that consist of complex systems of changing problems that interact with each other. I call such situations messes. Problems are abstractions extracted from messes by analysis; they are to messes as atoms are to tables and chairs.' Individual problems may be 'solved'. But if they are components of a mess, the solutions to individual problems cannot be added, since those solutions will interact. Problems may be solved; messes need to be managed. If we insist on the solution mode, analysts will be relegated to those relatively minor problems which are nearly independent, while messes go inadequately managed (Ackoff, 1981).

Rittel ... wicked versus tame problems

For Rittel, a 'tame' problem is one which can be specified, in a form agreed by the relevant parties, ahead of the analysis, and which does not change during the analysis. For a 'wicked' problem by contrast, there are many alternative types and levels of explanation of the phenomena of concern, and the type of explanation selected determines the nature of the solution. Alternative solutions are therefore not true or false, but good or bad. These judgements of worth must be made not by the analyst (who has no relevant expertise or standing in the matter) but by the interested parties themselves. According to Rittel 'the methods of Operations Research ... become operational ... only *after* the most important decisions have already been made, i.e. after the [wicked] problem has already been tamed' (Rittel and Webber, 1973).

Schon ... swamp versus high ground

Schon (1987) captures the dilemma of how good analysis can be carried out in messy, wicked situations via a vivid metaphor:

In the swampy lowland, messy, confusing problems defy technical solution. The irony of this situation is that the problems of the high ground tend to be relatively unimportant

to individuals or society at large, however great their technical interest may be, while in the swamp lie the problems of greatest human concern. The practitioner must choose. Shall he [*sic*] remain on the high ground where he can solve relatively unimportant problems according to prevailing standards of rigour, or shall he descend to the swamp of important problems and non-rigorous inquiry?

Ravetz ... practical versus technical problems

Technical problems are those for which at the inception of the study there exists a clearly specified function to be performed, for which a best means can be sought by experts. For a practical problem, by contrast, there will exist (at most) some general statement of a purpose to be achieved. The output of any study here should be, not a specification of optimal means, but an argument in favour of accepting a particular definition of the problem, together with its implication for the corresponding means of solution to be adopted. Practical problems, therefore, cannot be solved by technical or analytic expertise alone. This expertise must interact with judgement as to the cogency of arguments among diverse stakeholders (Ravetz, 1971).

(Source: Rosenhead and Mingers, 2001, pp. 4–6)

Juran and Gryna (1980) categorise problems in a rather different way that cross-cuts the distinctions offered in Box 1.2. They draw a distinction between sporadic and chronic. For these authors a sporadic problem is a sudden adverse change in the status quo, which is remedied by restoring the status quo. Where quality problems are concerned they link sporadic to the idea of control: sporadic problems arise when any process moves out of control, and they can be tackled by adjusting either the inputs to the process or, less frequently, the process itself. A chronic problem, on the other hand, is an adverse situation that has existed for a long time and is remedied by changing the status quo. Poor air quality near a busy road due to traffic emissions, low bathing-water quality due to the discharge of raw sewage, and erratic timekeeping by service engineers would be typical chronic quality problems.

Just as there are different categories of problem, so there are different types of solution. Ackoff (1978) draws a distinction between problems being solved, resolved and dissolved:

> A problem is said to be *solved* when the decision maker selects those values of the controlled variables which *maximize* the value of the outcome; that is, when he [*sic*] has optimized. If he selects

> values of the controlled variables that do not maximize the value of the outcome but produce an outcome that is good enough, he has *resolved* the problem by satisficing. There is a third possibility: he may *dissolve* the problem. This is accomplished by changing his values so that the choices available are no longer meaningful. For example, the problem of selecting a new car may be dissolved by deciding that the use of public transportation is better than driving oneself.
>
> (Ackoff, 1978, p. 13)

He uses the following example to emphasise the differences between these three:

> The differences between these approaches [are] illustrated by the following case. A large city in Europe uses double-decker buses for public transportation. Each bus has a driver and a conductor. The driver is seated in a compartment separated from the passengers. The closer the driver keeps to schedule, the more he [*sic*] is paid. The conductor collects zoned fares from boarding passengers, issues receipts, collects these receipts from disembarking passengers, and checks them to see that the correct fare has been paid. He also signals the driver when the bus is ready to move on after stopping to receive or discharge passengers. Undercover inspectors ride the buses periodically to determine whether conductors collect all the fares and check all the receipts. The fewer misses they observe the more the conductors are paid.
>
> To avoid delays during rush hours, conductors usually let passengers board without collecting their fares and try to collect them between stops. Because of crowded conditions on the bus they cannot always return to the entrance in time to signal the driver to move on. This causes delays that are costly to the driver. As a result hostility has grown between drivers and conductors which has resulted in a number of violent episodes.
>
> Management of the system first tried to ignore the problem, hoping that if it were left alone it would absolve itself. This effort at absolution did not work; the situation got worse.
>
> Management then tried to resolve the problem by proposing a return to an earlier state by eliminating incentive payments and accepting less on-schedule performance. The drivers and the conductors rejected this proposal because it would have reduced their earnings.
>
> Next management tried to solve the problem by having the drivers and conductors on each bus share equally the sum of the incentive payments due each. This proposal was also rejected by drivers and conductors; they were opposed to cooperating in any way.
>
> Finally, a problem dissolver was employed by management to deal with the situation. Instead of trying to compromise the conflicting

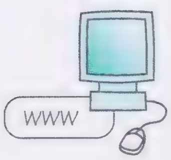
> interests of the drivers and conductors, he decided to take a broader view of the system. He found that during rush hours there were more buses in operation than there were stops in the system. Therefore, at his suggestion, conductors were moved off the buses at peak hours and placed at the stops. This reduced the number of conductors required at peak hours and made it possible to improve the distribution of their working hours. Under the new system conductors collected fares during peak hours from people waiting for buses and were always at the rear entrance to signal drivers to move on. At off-peak hours, when the number of buses in operation was fewer than the number of stops, conductors returned to the buses.
>
> The problem was dissolved.
>
> (Ackoff, 1999, pp. 115–16)

Ackoff also issues an important warning that is applicable to all types of solution – problems can become unsolved:

> Few problems, once solved, stay that way. Changing conditions tend to *un*solve problems that previously have been solved.
>
> [...]
>
> Because problems do not stay solved and their solutions create new problems, a problem-solving system requires more than the ability to maintain or control solutions that have been implemented and an ability to identify problems when or before they arise.
>
> (Ackoff, 1978, pp. 189, 190)

He is, of course, assuming that the problem was genuinely solved in the first place. As the American writer H. L. Mencken (sometimes known as 'the sage of Baltimore') famously said: 'there is always a well-known solution to every human problem – neat, plausible, and wrong' (Mencken, 1920, p. 158).

ACTIVITY 1.2 ....



Continue your analysis of the two problems that you worked with in Activity 1.1. Look at each of the perspectives in Box 1.2 and try to identify any insights they add to your understanding of the two problems. Assess how useful the various dichotomies are in your cases. ●

2.2 Categories of improvement

Distinctions are also drawn between different types of improvement. The category of improvement that is best known, largely due to its link with modern quality management, is continuous improvement. Ishikawa pioneered the quality circle movement in Japan in the 1950s as a mechanism for delivering continuous improvement, and by the 1980s the so-called 'quality

gurus' were stressing the need for the philosophy of continuous improvement to be taken up in the West. Juran, for example (see Juran and Gryna, 1980), taught that quality management was made up of three prongs: quality control, quality improvement and quality planning. In his terminology, quality improvement was about finding ways to do better than 'the standard'. Feigenbaum (1983) described quality as 'a way of managing the organisation' and control as a management tool with four steps:

a. Setting quality standards
b. Appraising conformance to those standards
c. Acting when the standards are exceeded
d. Planning for improvements in the standards.

(Feigenbaum, 1983, pp. 823–4)

However, the 'quality guru' who perhaps placed most emphasis on the need for continuous improvement is Deming. Look at Deming's 14-point plan for the achievement of Total Quality Management (TQM) in Box 1.3 and note the extent to which he emphasises the need for improvement over and over again.

BOX 1.3 DEMING'S 14 POINTS

(Note: the headings given to the 14 points by different authors vary, as does the order in which they are listed. I have culled those below from a number of sources but they retain Deming's own ordering.)

Point 1. Create constancy of purpose towards improvement of the product and service in order to become competitive, stay in business, and provide jobs.

Point 2. Adopt the new philosophy: we are in a new economic age. We no longer need live with commonly accepted levels of delay, mistake, defective material and defective manufacture.

Point 3. Cease dependence on mass inspection; require, instead, statistical evidence that quality is built in.

Point 4. Improve the quality of incoming materials. End the practice of awarding business on the basis of price alone. Instead, depend on meaningful measures of quality, together with price.

Point 5. Find the problems; constantly improve the system of production and service. There should be continual reduction of waste and continual improvement of quality in every activity in order to yield a continual rise in productivity and a decrease in costs.

Point 6. Institute modern methods of training and education for all. Modern methods of on-the-job training use control charts to determine whether a worker has been properly trained and is able to perform the job correctly. Statistical methods must be used to discover when training is complete.

Point 7. Institute modern methods of supervision: the emphasis of production supervisors must be to help people to do a better job. Improvement of quality will automatically improve productivity. Management must prepare to take immediate action on reports from supervisors concerning problems such as inherited defects, lack of maintenance of machines, poor tools or fuzzy operational definitions.

Point 8. Fear is a barrier to improvement, so drive out fear by encouraging effective two-way communication and other mechanisms that will enable 'everybody to be part of change, and to belong to it'.

Point 9. Break down barriers between departments and staff areas. People in different areas such as research, design, sales, administration and production must work in teams to tackle problems that may be encountered with products or service.

Point 10. Eliminate the use of slogans, posters and exhortations aimed at the workforce, demanding zero defects and new levels of productivity without providing methods. Such exhortations only create adversarial relationships; the bulk of the causes of low quality and low productivity belong to the system, and thus lie beyond the power of the workforce.

Point 11. Eliminate work standards that prescribe numerical quotas for the workforce and numerical goals for people in management. Substitute aids and helpful leadership; use statistical methods for continual improvement of quality and productivity.

Point 12. Remove the barriers that rob hourly workers, and people in management, of their right to pride in their work. This implies, among other things, abolition of the annual merit rating (appraisal of performance) and of management by objectives. Again, the responsibility of managers and supervisors must be changed from sheer numbers to quality.

Point 13. Institute a vigorous programme of education and encourage self-improvement for everyone. What an organisation needs is not just good people; it needs people who are improving with education. Advances in competitive position will have their roots in knowledge.

Point 14. Top management's permanent commitment to ever-improving quality and productivity must be clearly defined and a management structure created that will continually take action to follow the preceding 13 points.

Continuous improvement is often referred to by the Japanese word *kaizen*, especially when it is realised by making better use of existing resources. Eliminating waste through *kaizen* in Japanese companies involves an attempt to harness the mental as well as the manual skills of shop-floor workers. Individuals are encouraged to make suggestions – through quality circles, suggestion schemes and so on – about how savings can be achieved. The improvements sought through *kaizen* include maximum utilisation of labour through the elimination of unnecessary movements and idle time. When improvements have been incorporated into a job and the task redesigned,

standardisation is then carried out. For example, at the Toyota car plant in Japan each operator must be able to perform a standard repeatable sequence of operations in a given time (the time taken to produce a component or vehicle) with a specified quantity of parts to work on. This is recorded on paper and displayed at the worksite as a visual control. The standard then acts as a benchmark for further improvements. Other plants use a traffic light system (known as andon lights) to indicate how smoothly their lines are running. Green indicates that all workers are within cycle times, red that the line has stopped, and amber that there are some problems.

Two other categories of improvement that are familiar in Japan are *kaikaku* (radical improvement) and *kairoy* (improvement achieved by innovation and investment in new plant or systems). In the West, radical improvement is also known as breakthrough. The term 'breakthrough' was introduced by Juran, when he was working with the Japanese in the 1950s and 1960s, to refer to the solution of chronic quality problems. He defined a breakthrough sequence for solving chronic quality problems as follows:

1 Convince others that a breakthrough is needed – convince those responsible that a change in quality level is desirable and feasible.

2 Identify the vital few projects – determine which quality problem areas are most important.

3 Organise for breakthrough in knowledge – define the organisational mechanisms for obtaining missing knowledge.

4 Conduct the analysis – collect and analyse the facts that are required and recommend the action needed.

5 Determine the effect of proposed changes on the people involved and find ways to overcome the resistance to change.

6 Take action to institute the changes.

7 Institute controls to hold the new level.

(Juran, 1964)

There are two types of innovation: incremental and radical. Improvement achieved by innovation can therefore range from a change that is indistinguishable from one attributed to continuous improvement to one that is, in a sense, beyond breakthrough because its consequences are so dramatic.

A major survey (Leach et al., 2001) looking at innovation in UK companies was able to identify the attributes of organisations that were associated with successful innovation. These are shown in Box 1.4. The attributes are almost exactly those you would expect to find in organisations that are successful at other forms of improvement.

BOX 1.4 ATTRIBUTES OF MORE SUCCESSFUL MAJOR INNOVATIONS

Work–environment attributes

Reflecting the factors associated with innovation more generally ..., the degree of success of particular innovations was found to be greater in organisations that:

- Benchmarked their operations more frequently against other organisations.
- Received feedback more frequently from customers or clients about products or services.
- Captured more ideas from non-management employees and gave greater feedback about their ideas.
- Operated a formal communication system for sharing strategic information with all employees.

Process attributes

More successful innovations were found in organisations that:

- Conducted extensive internal (within organisation) and external (with other organisations) discussion and negotiation prior to idea implementation.

Innovation attributes

More beneficial outcomes were associated with innovations that:

- Changed significantly the way in which the organisation operates.
- Affected most of the organisation.

Factors *not* associated with more successful innovations included:

- The level of human or financial resources invested in the innovation.
- Extent of departure from what the organisation had done before.
- Extent of departure from what any other organisation had done before.
- How risky the innovation was for the organisation.

(Source: Leach et al., 2001)

Now read Offprint 1

3 APPROACHES TO PROBLEM SOLVING AND IMPROVEMENT

In the Introduction I emphasised the advantages of problem solving and improvement based on 'systematic rational thought supported by relevant information' (Grünig and Kühn, 2005, p. 8). The best way of making sure you are being systematic is to use an approach or method. Problem-solving and improvement approaches are organised, and incorporate procedures, to ensure that a pattern is followed. The path that is set, even if it allows multiple branching and variety within it, forces the user of the approach to confront difficult but important issues rather than selectively ignore them. There is a further advantage of using an approach or method: it allows someone else to examine the way recommendations have been generated and so form an opinion on whether the recommendations warrant confidence and are likely to be successful.

3.1 Different types of approach

One of the most fundamental differences between approaches is whether they are reductionist or holistic. The reductionist thinking process was largely designed to solve scientific problems and to guide scientific research. Drawing on the writings of John Platt, Waddington (1977) describes the scientific method as consisting of the following steps:

> (1) devising alternative hypotheses;
>
> (2) devising a crucial experiment (or several of them) with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
>
> (3) carrying out the experiment so as to get a clean result; and recycling the procedure, making subhypotheses or sequential hypotheses to define the possibilities that remain;
>
> and so on.
>
> (Platt, 1964, cited in Waddington, 1977, pp. 118–19)

This method, which Waddington, citing Platt, argues should be called the 'method of strong inference' has spread beyond scientific enquiry into many other areas. In essence, where problem solving is concerned, the scientific method has become bound up with the reductionist approach. This is characterised by taking a problem, decomposing it into individual sub-problems and discovering how to solve the individual sub-problems. The belief is that this will allow root causes of problems to be identified and dealt with, and that by building up a detailed understanding of every part it will be possible to remove each root cause in turn and by so doing improve the whole.

A holistic approach tries to treat a problem or a situation as a whole and regards the interactions between components as just as important as the

components themselves when carrying out investigations and generating recommendations. Any changes are assessed in terms of their effect on the operation of the whole rather than on specific aspects.

Table 1.2 shows a comparison of holistic and reductionist problem-solving approaches.

Table 1.2 Comparison of holistic and reductionist approaches

Holistic solution creation	**Reductionist problem solving**
Employs many mental models: intuitive, analytic, creative	Employs rational, empirical thought process
Future oriented; focuses on creating solutions	Past oriented; focuses on solving each problem
People centered	Fact centered
Seeks out broad context in which to understand a problem and its potential solutions	Limits context to the problem itself
Aims to find unique, novel ideas that provide the basis for a living solution that can endure and change over time	Aims to find a single, immediate solution that 'fixes' the problem
Recognizes that all information is soft	Emphasizes only hard data
Initially treats each problem situation as unique	Seeks similarities with other problems
Puts solutions in a system framework, recognizing interdependencies with other systems	Specifies changes only in terms of the parts of the problem

(Source: Nadler and Chandon, 2004, p. 15)

ACTIVITY 1.3

I would suggest that, in order to emphasise the differences between holism and reductionism, the authors have polarised the two across all rows of the table. In your opinion which row identifies the essential difference between holism and reductionism? Try to find three examples in the table where the dichotomy that has been set up has been exaggerated or is false. ●

Another way of distinguishing between approaches is whether they are heuristic or analytic. The standard definition of a heuristic is shown in Box 1.5.

BOX 1.5 DEFINITION OF A HEURISTIC

> A heuristic ... is a rule of thumb, strategy, trick, simplification, or any other kind of device which drastically limits search for solutions in large problem spaces. Heuristics do not guarantee optimal solutions; in fact they do not guarantee any solution at all; *all that can be said for a useful heuristic is that it offers solutions which are good enough most of the time.*
>
> [...]
>
> A useful rule of thumb used by human beings in most of their problem-solving is this: Attack a new problem by methods that have solved similar problems in the past. The criteria for 'similarity' may themselves be heuristic. If the environment is in a kind of steady state with respect to problem types, this heuristic may be very useful. In environments demanding a high degree of innovative problem-solving, this heuristic will hinder rather than facilitate problem-solving.
>
> ... Two general-purpose heuristic problem-solving methods commonly employed in human reasoning are *means-end analysis* and *planning*. In means-end analysis, an initial problem state is transformed into a target state by selecting and applying operations which, step by step, reduce the difference between the states. In the planning method, a simplified statement of the original problem is constructed, and means-end analysis is applied to this new, simpler problem. The result is a set of plans ..., hopefully one of which will work, *i.e.*, solve the original problem.
>
> (Source: Feigenbaum and Feldman, 1963, pp. 6, 7)

Analytic problem solving is based on the use of techniques to conduct rigorous analysis in order to try to find optimal solutions. Most analytic procedures have significant application restrictions. Some, for example, are recommended for use only when: all aspects of the problem can be expressed in terms of quantifiable variables; it is possible to know in advance what would constitute a solution; and selection between possible solutions can be achieved by the application of clear criteria. (These requirements are shown in diagrammatic form, overleaf, in Figure 1.2.) Usually, such criteria must cover four key areas:

1 content, i.e. what will be achieved
2 level of attainment
3 how long the solution will remain valid
4 the scope of the solution.

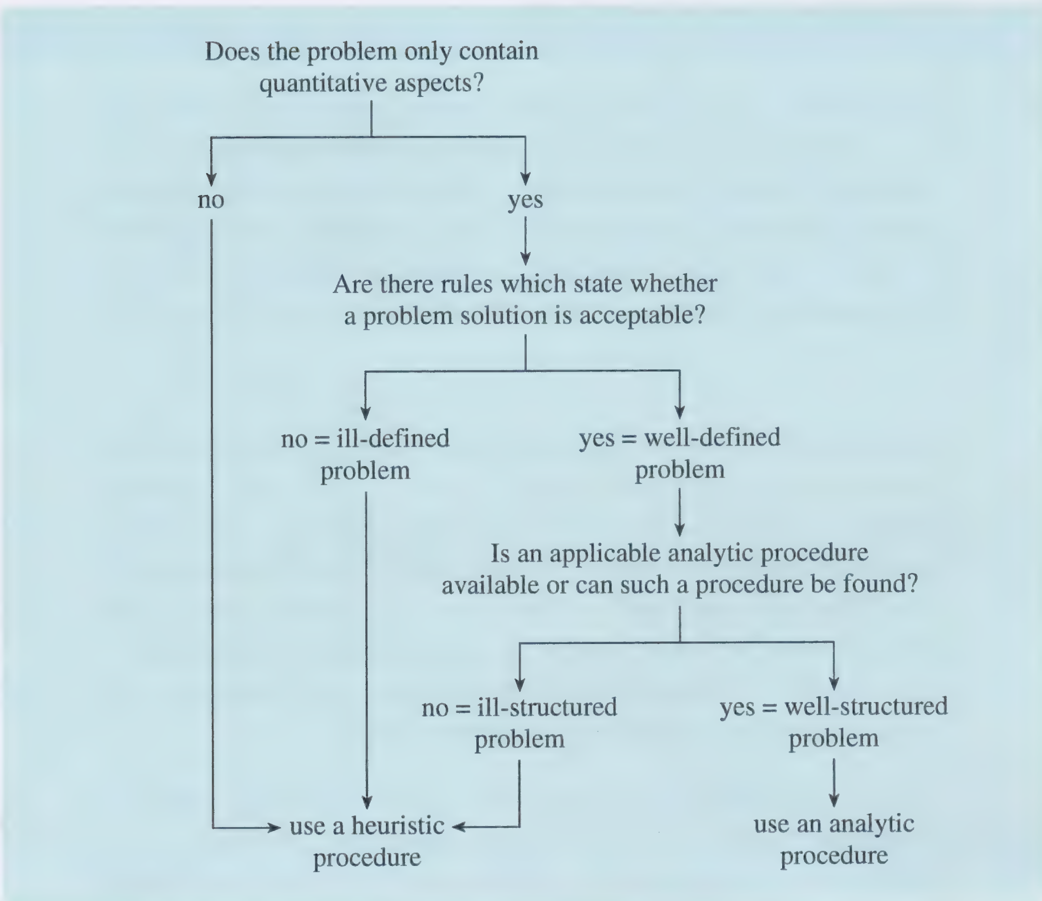


Figure 1.2 The three requirements for using an analytic procedure (Source: Grünig and Kühn, 2005, p. 51)

Of course it is not always that straightforward. Most problem-solving and improvement exercises have a number of different objectives, some aspects are not quantifiable, and there is conflict within the criteria for selecting solutions. There is a lot of truth in the saying that you can achieve any two goals out of cheap, fast and effective but you cannot have all three. In many situations it is therefore necessary for the people involved to agree a set of objectives and prioritise them if necessary. The word 'set' is important here. The most effective problem solving and improvement optimises the whole, but there is always a temptation to sub-optimise and settle for a satisfactory outcome rather than the best one. In order to look across the whole, it is essential to identify criteria that cover as many objectives as possible so that judgements can be made across them.

Grünig and Kühn sum up the relative merits of heuristic versus analytic as follows:

> The essential advantage of heuristics in comparison to analytic procedures lies in the almost total absence of formal application restrictions and in their relatively low application costs. The disadvantages are the absence of any guarantee of a solution and,

where a solution is found, the lack of guarantee that it is the optimal solution.

(Grünig and Kühn, 2005, p. 48)

3.2 Generic problem-solving and improvement methods

Block 4 of this course contains a number of different methods and approaches to problem solving and improvement. Many of those approaches have terminology, and perhaps one or two techniques, that are unique to them, but by and large most approaches to problem solving and improvement draw on the same or similar techniques even if they sometimes call them different names or use them for different purposes. Tools and techniques are covered in Block 3, so by the time you reach Block 4 you will have studied most of the techniques included in the course. In order to provide a framework for understanding those techniques and starting to use them, I shall look at three generic problem-solving and improvement methods now. They are based on three different metaphors for problem solving:

1 a learning cycle
2 a journey
3 a search

Metaphor 1: a learning cycle

Many problem-solving and improvement methods encapsulate a cycle of activity that leads from some initial alerting event through a thought process, investigation and analysis to action, and then round the cycle again and again to try to achieve further improvement. Perhaps the best known of these is the P–D–C–A (plan–do–check–act) cycle shown in Figure 1.3. This is often also called the Deming wheel or the

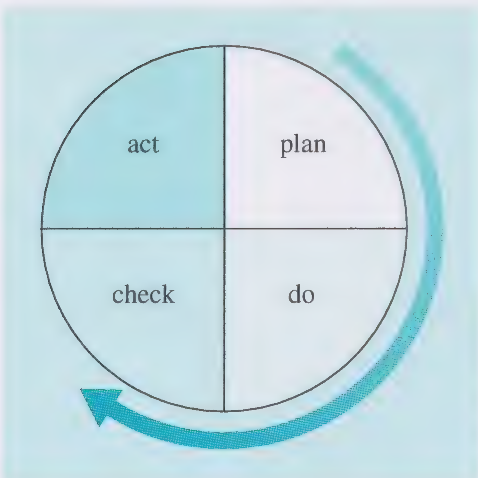


Figure 1.3 The P–D–C–A cycle

Shewhart cycle and sometimes has an E added to the front of the abbreviation to give E–P–D–C–A:

- Evaluate and define objectives.
- Plan to achieve those objectives fully.
- Do (implementation of plans).
- Check that objectives are being met.
- Act if corrective action needed.

Many users have customised this cycle. For instance, Cameron-Jones (1985) has developed a self-improvement version for use by teachers where the stages are:

1 Choose a general idea and analyse it; for example, the different amounts of time spent individually with different pupils, and whether these differences are justifiable or should be changed/improved.
2 Make a reconnaissance of your present teaching.
3 Devise an action plan.
4 Take the action you have planned.
5 Monitor it.
6 Reflect.

Oakland and Marosszeky (2006) call their version the DRIVE model for continuous improvement where D stands for define the problem, R for review the information, I for investigate the problem, V for verify the solution and E for execute the change. Here too it is going round and round the cycle that delivers continuous improvement.

Conceptually, methods such as these echo the phases of an action learning cycle of the type set out in Figure 1.4 and described more fully in Table 1.3.

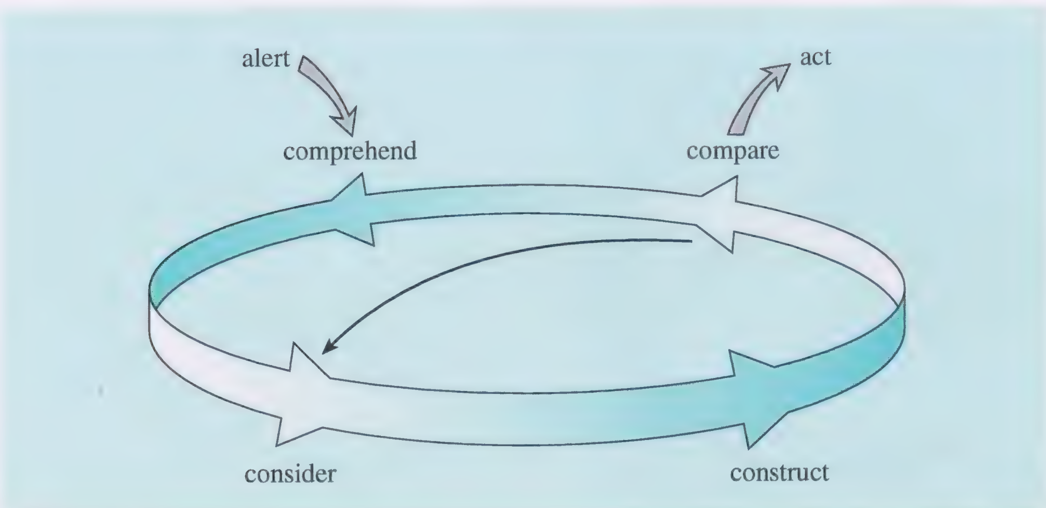


Figure 1.4 An action learning cycle

Table 1.3 The elements of an action learning cycle

Cycle element	Description
Alert from the environment	The act of initiating the inquiry. This could be a decision by the problem solver, or a decision within an organisation to devote resources to a specific task. Whatever form this event takes, its effect is to cause resources to be applied to the analysis of some situation.
Comprehend the situation now	This is action, or more properly a series of actions, through which an understanding of a situation is acquired. It can include interviews, discussions, workshops and surveys of perceptions, attitudes and knowledge. Acquiring this understanding may be facilitated by a wide range of techniques and at some stage it will be useful to bring together the data collected so that it can be considered as a whole.
Consider what to do	This is the act of planning the specific activities to be undertaken in this analysis.
Construct models	This is the act of creating models considered appropriate to this situation in order to generate a range of possible futures; but do beware – this is a simple action-process to define but a complex one to carry out.
Compare models with 'reality'	In this element the models created are compared with what is known about the situation in order to bring about some end-activity to this cycle of analysis. A range of end-activities are feasible. Maybe some action can be proposed or discussed. In some situations a clear recommendation for one particular option (out of several) can be made. In some circumstances the best move is to iterate within the current analysis to clarify or expand the work. In other cases it may be appropriate to move into a new cycle of analysis. Whichever of these (or combination of them) is the case, there is then the need to articulate a plan to move on.
Act	This culminates a cycle by carrying out the planned end-activity arising from the comparison. This could result in change and closure of the analysis activity, at least for the time being, or continuance in another cycle of analysis.

Metaphor 2: a journey

The second metaphor I shall look at is problem solving and improvement as a journey. Using this metaphor the purpose of the journey is to move from where you are now (the current state of affairs) to where you want to be (a more desirable state of affairs). This may mean removing or containing something undesirable in the current state so that its effects are no longer experienced in the final state, and/or acquiring something desirable that is absent at the start so that it becomes available at the finish.

For any journey there are usually a number of different ways of getting from start to finish and it is likely that each route and mode of transport will have some advantages and some disadvantages. It is therefore necessary to have a way of selecting the best route. Once the selection has taken place more detailed planning can start and all that is left is to undertake the journey and monitor progress as the journey unfolds.

Figure 1.5 shows a method based on this metaphor. Use of the method begins when it is thought there is a problem to be solved or an opportunity to be exploited. Stage 1 is discovering more about where you are now. Knowing that there is a problem or opportunity is not the same as having a clear idea what it is, so one of the aims of this stage is to define the problem or opportunity sufficiently well to allow analysis to begin.

Stage 2 is identifying a set of objectives (where you want to get to) and any constraints that can restrict the choice of objectives or prevent their being reached. Objectives must address the problem to be solved or the

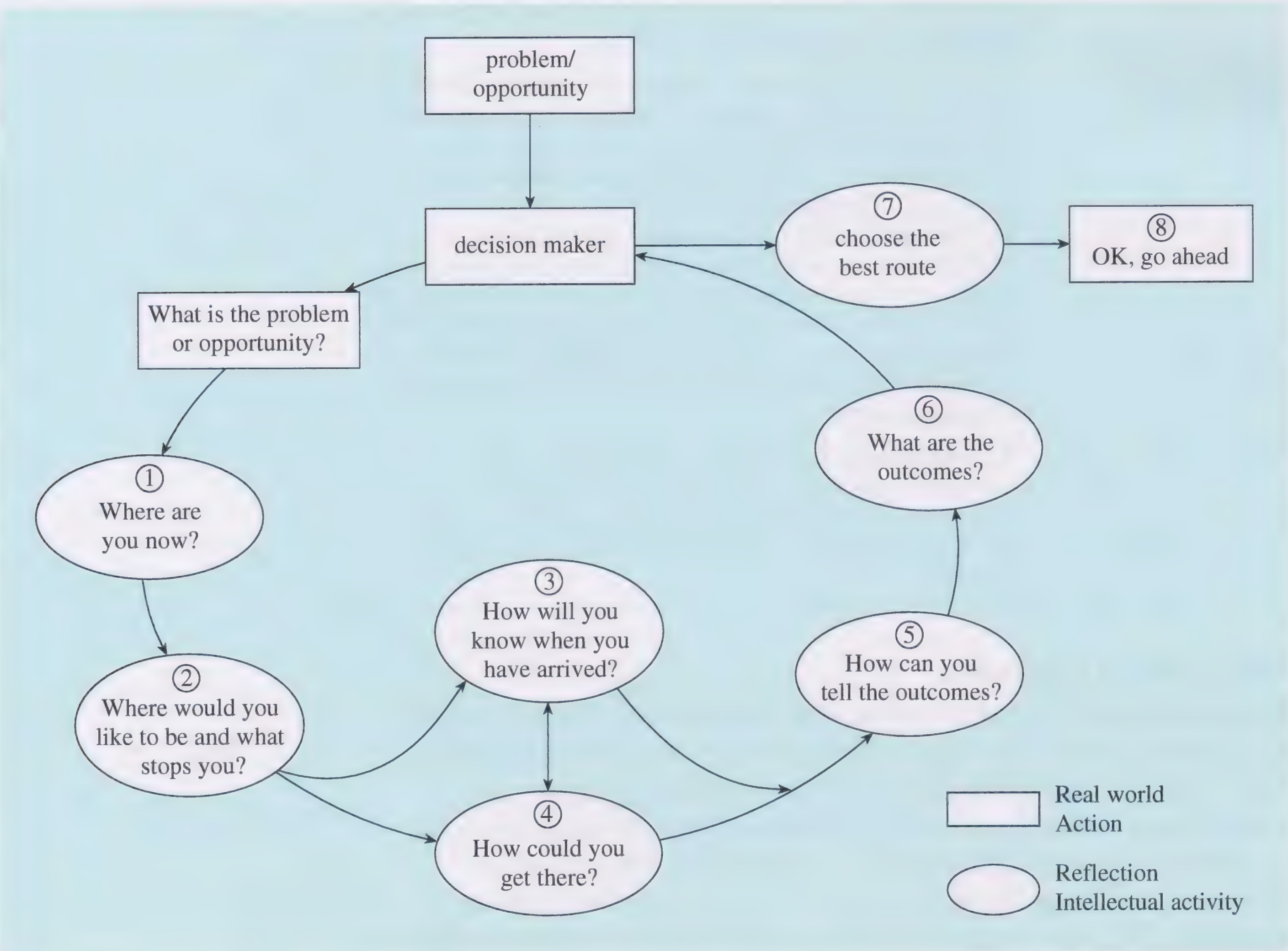


Figure 1.5 The journey method of problem solving and improvement

opportunity to create something new and embody what needs to be done in order to deliver solutions. Some are likely to be quantitative (for example, increase yield by 15 per cent) but others will be qualitative (for example, improve cooperation between departments) and it will be much more difficult to know when the latter have been met. Sometimes a constraint can be negotiated away (for example, 'within a budget of £25 000' can be circumvented by increasing the budget) and sometimes it can be dealt with by bringing it into the scope of the problem-solving or improvement project. Again, one of the biggest difficulties is conflict between objectives.

Stage 3 is the formulation of measures of performance, so it is necessary to:

1 state the measures/forms of assessment that will be used
2 establish target levels
3 set timescales over which they should be met
4 establish acceptable levels of error for the target values.

Measures of performance are closely related to objectives and are the means by which success in meeting objectives is measured. Some objectives, such as profitability, are easily measured, although even here there would be scope for controversy over the rates of depreciation applied to capital equipment or the value placed on stocks and work in progress. Other objectives, such as improving labour relations or obtaining an enhanced reputation for good quality in the marketplace, are much more difficult to measure. In general, therefore, when setting an objective, it is important to consider at the same time a means of measuring the system's success in meeting that objective.

All the objectives identified in stage 2 need to be covered by the measures but this does not mean that each objective must have a separate measure. Indeed, measures that look across the whole set are more able to identify optimal solutions.

ACTIVITY 1.4

Suggest three measures that could be used to examine the performance of a purchasing department. ●

Stage 4 involves generation of routes to objectives so it is where potential solutions are developed. Some of these may be based on 'blue-sky thinking', with novel solutions being developed from scratch, but others will involve incremental changes derived from careful analysis of the current situation. In both cases it is necessary to draw on techniques you will meet in later blocks.

The next task (stages 5 and 6) is to assess, in terms of the measures of performance, the likely outcomes of taking each of the routes to the

objectives. It is very unlikely that one route is best for all objectives. Probably the biggest challenge of using the journey method in a complex situation is deciding which route is the best overall (stage 7) but, once the recommendation has been made, an action plan can be drawn up before implementation begins.

The Value for Money Team at the National Audit Office sets out a very useful set of questions (the 'six Ws') that recommendations need to address if they are to be useful:

> What needs to be done?
>
> Why does it need to be done?
>
> Where does it need to be done?
>
> When does it need to be done?
>
> How is it to be done?
>
> Who is to do it?
>
> (National Audit Office, 2002, p. 13)

Metaphor 3: a search

The metaphor of problem solving as a search is most closely associated with operational research and artificial intelligence but it shares many features with the previous metaphor. A representation of a problem and a description of an ideal solution are formulated, and the task is to search for possible solutions and then select one that is equal to or close to the ideal solution.

Muñoz-Seca and Riverola (2004) describe the process thus:

> The solver looks among a series of possibilities for a solution that is pleasing to him [*sic*]. During the process, he redefines the problem's structure, broadening or narrowing horizons and defining alternatives and possibilities.
>
> (Muñoz-Seca and Riverola, 2004, p. 10)

They then go on to identify the functional components of problem solving conceptualised as a search process:

1 Goals and constraints.

2 The problem's state. This is 'a description of all the elements of its history that are relevant for the future' (p. 13).

3 Measures of distance. These show 'how far away we are from obtaining the "solution" to our problem' (p. 15).

4 A set of transformations. These 'can be applied to the current state, to make it evolve towards new states. To continue progressing in the solution process, the problem solver selects one from among the series of possible transformations (that he knows)' (p. 15).

5 Search mechanism. This mechanism is used to search for the solution. 'Generally speaking, the search mechanism has three basic components: the transformation selection mechanism; the backtracking mechanism; and the mechanism for selecting the solution to be explored next' (p. 17).

6 Knowledge state, heuristic rules and models. The search process is used to generate learning in the way shown in Figure 1.6.

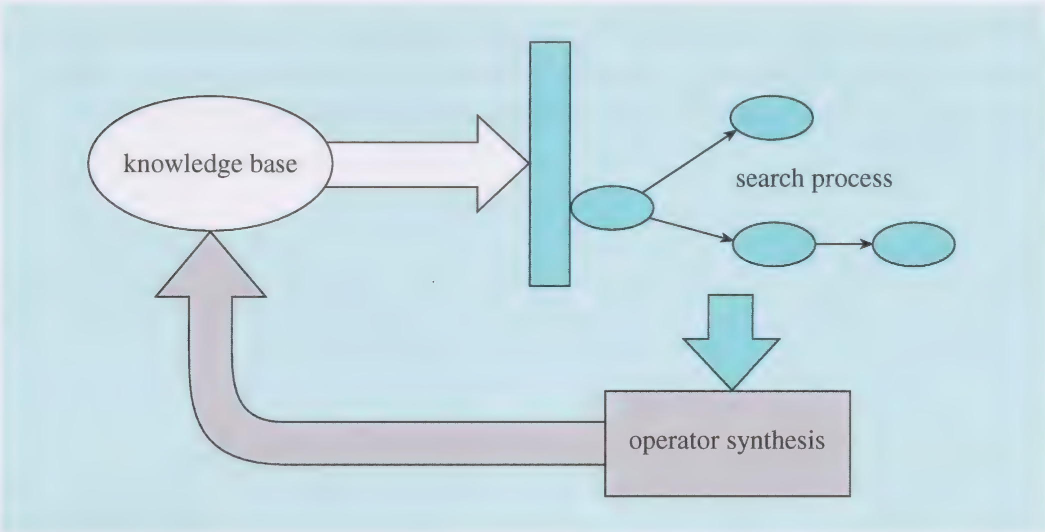


Figure 1.6 The search process and learning (Source: Muñoz-Seca and Riverola, 2004, p. 18)

4 CONCLUSION

This block has set the scene for the course by looking at the background to problem solving and improvement, and identifying some of the key terminology. It has also introduced different types of problem-solving and improvement approach and introduced three generic approaches based on the metaphors of learning, a journey and a search.

The block is general and discursive and is designed to raise your awareness. The next block is radically different. It introduces statistical concepts and tools that can be used to identify, describe and understand problems and opportunities.

REFERENCES

Ackoff, R. L. (1978) *The Art of Problem Solving*, New York, Wiley.

Ackoff, R. L. (1979) 'The future of operational research is past', *Journal of Operational Research Society*, Vol. 30, pp. 93–104.

Ackoff, R. L. (1981) 'The art and science of mess management', *Interfaces*, Vol. 11, pp. 20–6.

Ackoff, R. L. (1999) *Ackoff's Best*, New York, Wiley.

Cameron-Jones, M. (1985) *Focus on Teaching: Project Report*, Edinburgh, Moray House College of Education.

Feigenbaum, A. V. (1983) *Total Quality Control*, New York, McGraw-Hill.

Feigenbaum, E. A. and Feldman, J. (eds) (1963) *Computers and Thought*, New York, McGraw-Hill.

Flood, R. L. and Jackson, M. C. (1991) *Creative Problem Solving*, Chichester, Wiley.

Grünig, R. and Kühn, R. (2005) *Successful Decision-making*, Berlin, Springer.

Juran, J. M. (1964) *Managerial Breakthrough*, New York, McGraw-Hill.

Juran, J. M. and Gryna, F. M. (1980) *Quality Planning and Analysis*, New York, McGraw-Hill.

Leach, D., Totterdell, P., Birdi, K., Clegg, C., Wood, S. and Wall, T. (2001) *Innovation at Work*, Sheffield, The Institute of Work Psychology, University of Sheffield, http://esrccoi.group.shef.ac.uk/worddocs/innovation.doc (accessed 11 February 2007).

Mencken, H. L. (1920) *Prejudices: Second Series*, New York, Knopf.

Muñoz-Seca, B. and Riverola, J. (2004) *Problem-Driven Management*, Basingstoke, Palgrave Macmillan.

Nadler, G. and Chandon, W. J. (2004) *Smart Questions*, San Francisco, Jossey-Bass.

National Audit Office (2002) *Writing Smart Recommendations*, National Audit Office, http://www.concordat.org.uk/_db/_documents/NAO_Smart_Recommendations.pdf (accessed 11 February 2007).

Oakland, J. and Marosszeky, M. (2006) *Total Quality in the Construction Supply Chain*, Oxford, Butterworth-Heinemann.

Platt, J. (1964) 'Strong inference', *Science*, Vol. 146, pp. 347–53.

Ravetz, J. R. (1971) *Scientific Knowledge and its Social Problems*, Oxford, Oxford University Press.

Rittel, H. W. J. and Webber, M. M. (1973) 'Dilemmas in a general theory of planning', *Policy Science*, Vol. 4, pp. 155–69.

Rosenhead, J. and Mingers, J. (2001) 'A new paradigm of analysis', in Rosenhead, J. and Mingers, J. (eds) (2001) *Rational Analysis for a Problematic World Revisited*, Chichester, Wiley, pp. 1–19.

Schon, D. A. (1987) *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*, San Francisco, Jossey-Bass.

Waddington, C. H. (1977) *Tools for Thought*, London, Jonathan Cape.

ANSWERS TO ACTIVITIES

Activity 1.3

In my view the bottom row of the table identifies the essential difference between holism and reductionism with the fourth row (broad context *v.* context limited to the problem itself) as a close second. Examples where the dichotomy that has been set up is exaggerated or false include:

- The use of empirical to describe only reductionist problem solving. Empirical means based on observation and I would suggest a holistic approach could also use an empirical thought process.
- The distinction between the use of soft information and hard data. I believe this has been exaggerated: 'all' and 'only' are too absolute in the comparison; 'much' and 'mainly' would be more accurate.
- People centred versus fact centred. This is also something of a caricature; it is not an inherent difference but more a comment based on inferences drawn about problem solvers who prefer holistic approaches compared with those who prefer reductionism.

Activity 1.4

Examples include: the quality of incoming goods measured in terms of the number of units that need to be returned to the supplier, the timeliness of deliveries, the unit costs of goods delivered, and the cost of operating the purchasing system.

ACKNOWLEDGEMENTS

Grateful acknowledgement is made to the following sources:

Text

Ackoff R.L., (1999), 'Problem Treatments', *Ackoff's Best: His Classic Writings on Management*, John Wiley & Sons; Flood, R.L. and Jackson, M.C. (1991), *Creative Problem Solving*, John Wiley & Sons.

Tables

Table 1.1: Flood, R.L. and Jackson, M.C. (1991), *Creative Problem Solving*, John Wiley & Sons; Table 1.2: Nadler, G. and Chandon, (2004), W.J. Smart Questions, Jossey-Bass.

Figures

Figure 1.2: Grunig R., and Kuhn R., (2005), 'Examples of the Different Types of Decision-making Procedures', *Successful Decision-making*, Springer-Verlag; Figure 1.6: Munoz-Seca B. and Riverola J., (2004), 'Search Process and Learning', *Problem-Driven Management*, Palgrave Macmillan.

Block 2 Statistics

CONTENTS

AIMS

The aims of Block 2 are to:

1 provide an introduction to the language of statistics and the basic statistical concepts and methods of analysis that underpin problem solving and improvement

2 introduce the statistical software package provided with the course and show how it can be used to produce a variety of graphical displays, to calculate numerical summaries and to conduct other statistical analyses.

LEARNING OUTCOMES

After studying Block 2 you should be able to:

1 present data graphically using a variety of methods

2 make the following calculations for a set of data: mean; median; mode; range; interquartile range; and standard deviation

3 understand the concept of relationships between variables

4 understand the concepts of probability and conditional probability, and be able to apply the rules of addition and multiplication of probabilities

5 demonstrate a working knowledge of the normal and the standardised normal distributions, the binomial distribution and the Poisson distribution, and be able to make a variety of calculations associated with them.

1 INTRODUCTION

There are two types of analysis: quantitative and qualitative. Many people seem to have a strong preference for conducting one or the other. At one extreme there are those who believe that 'if you cannot measure it you cannot manage it' and at the other there are those who run a mile from anything involving arithmetic and formulae. However, concentrating just on quantitative or just on qualitative is not at all helpful in problem solving and improvement; situations are almost always a combination of 'hard' and 'soft' features and thus call for quantitative and qualitative analysis in order to understand them fully. This course places equal emphasis on both the quantitative and the qualitative but it also recognises that your background and preferences mean that you will probably be more comfortable with one than the other. If your preference is for quantitative then most of the content of this block is unlikely to be new to you and its main benefit will be to introduce you to the statistical software package the course uses. However, if your preference is strongly qualitative this block is designed especially for you. The block introduces you to the language of statistics and to many of the fundamental concepts and methods of analysis that are used in many problem-solving and improvement techniques.

The computing content of the block will be new to everyone who has not already used the package that is supplied. It is important that you become familiar with the statistical software because there are parts of the course where use of it is essential (for example, study of statistical process control and design of experiments in Block 3) and to help you to do this there are a number of computer exercises associated with this block. They are flagged with an icon in the margin and set out in the Computer Exercises Booklet.

Although you need to become familiar with the statistical software package and use it in a few specific sections of the course, the amount of use you make of it apart from that is your choice. There are two reasons for leaving this decision up to you. The first is to recognise that practical considerations may dictate how you study. For example, you might be studying in a place or at a time when powering up your computer is not practicable. The second is to recognise your right to choose according to your personal preferences: a graph is just as valuable whether it is plotted using a ruler and pencil, a spreadsheet package, statistical software, or indeed any other way. I have chosen printed text as the primary vehicle for presenting the material in this block even though all the topics it deals with are available via the software. The reason for this is that understanding the concepts introduced in the block is very important, and moving straight to a statistical software package can mean you conduct an analysis without understanding what you are doing and what the results really mean. However, I would point out that there are very many advantages to using the software. It will make your work much easier and quicker and allow you to avoid (often tedious) calculations and physical

data manipulation. It also produces very professional-looking output that can easily be incorporated into assignments and your project report.

Now do Exercise 2.1 in the Computer Exercises Booklet.

Before I look at different ways in which data can be presented to enable people to understand their meaning more easily and quickly, I should like to address a few words to those of you who are either completely new to the subject of statistics or have given up on it once already. There is nothing very tricky about the statistics you will need to use while studying this course, provided that you don't fall prey to the two devices it uses to put people off. These are:

- Greek letters
- algebraic notation.

The first can easily be dealt with by learning a few new words such as that for the Greek equivalent to the letter S, which is written Σ, pronounced 'sigma', and used to indicate when a number of terms have to be added together. The algebraic notation is very useful as a form of shorthand and is well worth the amount of effort needed to get used to it. In algebraic notation, statisticians allow letters (such as x) to stand for numbers. For example, six numbers can be denoted as six different values of x as follows:

$$x_1, x_2, x_3, x_4, x_5 \text{ and } x_6$$

where the first number in the list of six is called x_1 and the final number is x_6. For the list:

6, 9, 1, 4, 5, 7

$x_3 = 1$

For the list:

8, 4, 8, 7, 3, 6

$x_3 = 8$

A short way of writing down the general form of the list is to present it as:

$$x_i;\ i = 1, 2, ..., 6$$

or even shorter still:

$$x_i;\ i = 1 \text{ to } n$$

where n represents the number of numbers.

I shall use Greek letters and algebraic notation that are the accepted language of statistics throughout this block, so please bear with them. The terms will be clearly defined when they are introduced, but you might also find it useful to refer to the glossary which is included towards the end of the block.

2 GRAPHICAL PRESENTATION OF DATA

Data can be classified into a number of different types. The main types are shown in Table 2.1. When choosing between the graphical methods you will meet in this section it is important to consider the type of data you want to represent because not all methods are suitable for all types of data.

Table 2.1 Types of data

Type of data	**Level of measurement**	**Examples**
Categorical	Nominal (no inherent order in categories)	Colour, diagnosis, product ordered by customer
	Ordinal (categories have inherent order)	Job grade, level of satisfaction with service
	Binary (2 categories – special case of nominal or ordinal above)	Gender
Quantitative	Discrete (usually whole numbers)	Size of household, number dispatched per day
	Continuous (can, in theory, take any value in a range, but usually recorded to a predetermined degree of precision)	Temperature, weight, length of time

Information that would be difficult to assimilate when presented as a string of numbers can be much easier to comprehend if presented graphically. For example, consider the set of numbers in Table 2.2, represented in statistics notation by:

x_i; $i = 1$ to n.

Table 2.2 Values for x_i; $i = 1$ to 10

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
3.0	4.0	3.6	3.0	3.8	2.3	3.5	5.0	3.5	4.0

Examining the data, you might want to know whether any trend existed in them, that is, whether the values were gradually increasing or decreasing. From the table you would be able to view the data only discretely – that is, one number at a time – and try to make a judgement on that basis. However, if the same numbers were presented in graphical form, as in Figure 2.1, you would be able to make a much more immediate decision.

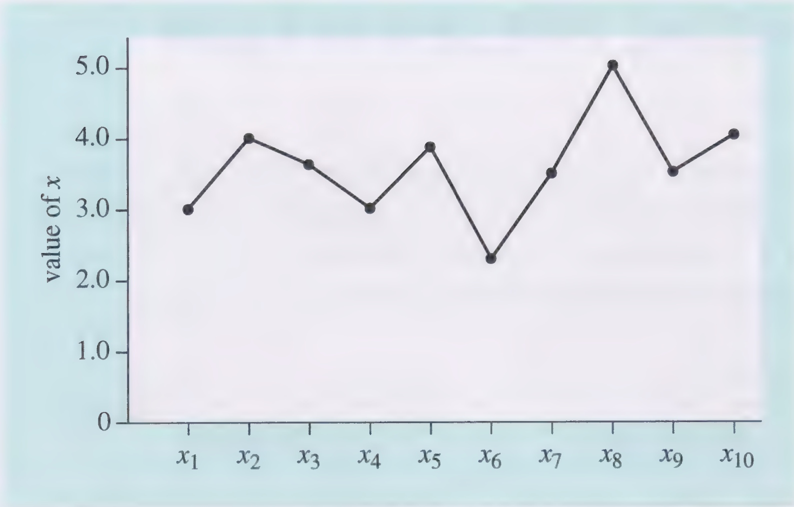


Figure 2.1 Graph for values of x_i; $i=1$ to 10 as given in Table 2.1

The reason for this is that graphical presentations tap intuitive abilities which stem from the combined powers of human vision and the brain. Trends or features that can be buried in a welter of numbers can be brought into sharp focus in a graph. Apart from line graphs such as that in Figure 2.1, there are numerous other ways of presenting data graphically. I am going to look at those which are commonly used to present information about quality.

2.1 Tally charts

A *tally chart*, as its name suggests, is a chart that is used to keep a 'tally' or record of certain objects, observations or occurrences. For instance, suppose that you were concerned about your organisation's energy use and wanted to see how many computers had been left on over the weekend, so you made a tour of the building one Saturday afternoon to check. One efficient way of recording the results of your investigation would be to draw up a chart similar to the one in Figure 2.2. For every computer you found that had been shut down ('off') you would enter a line or a 1 in one row, indicating every fifth machine by a horizontal or diagonal bar through the four previous lines, and entering the sixth machine in a new column further along the row. In the other row you would follow the same procedure for those not shut down ('on').

									class frequency
off	𝍸	𝍸	𝍸	𝍸	||				22
on	𝍸	|							6

Figure 2.2 Tally chart showing machines 'off' and 'on'

When you had gone through the entire building you would have completed a tally chart giving you an instant picture of the relative numbers of machines that were on and off. A tally chart typically consists of a series of rows representing different categories or classes. In the example there are two rows but any number of categories might be used. The total counts for each class or category are known as the frequency. Viewed together, the class frequencies are known as the frequency distribution.

ACTIVITY 2.1

Below are the results from a test carried out on a sample of 50 light bulbs. The figures indicate the length of time (to the nearest hour) that each light bulb functioned continuously before burning out. Construct a tally chart from these data using the following classes: 0–249 hours; 250–499 hours; 500–749 hours; and 750–999 hours.

850	790	760	770	280	950	901	920	880	920
973	501	730	899	600	898	450	911	50	903
965	990	902	913	956	897	972	898	876	877
900	888	520	926	950	925	702	803	955	811
820	901	950	942	650	873	984	909	550	909

2.2 Bar charts and histograms

A bar chart is used for presenting categorical data. As you can see in Figure 2.3, it is a type of graph consisting of two axes and a series of vertical bars or columns. The horizontal axis, or x-axis, shows the categories and the vertical axis, or y-axis, indicates the frequency of the respective category. The y-axis usually begins at zero. The example in Figure 2.3, which relates to the tally chart in Figure 2.2, has just two categories, off and on. Note that the bars do not touch and the width of each class or column is the same; only the height varies.

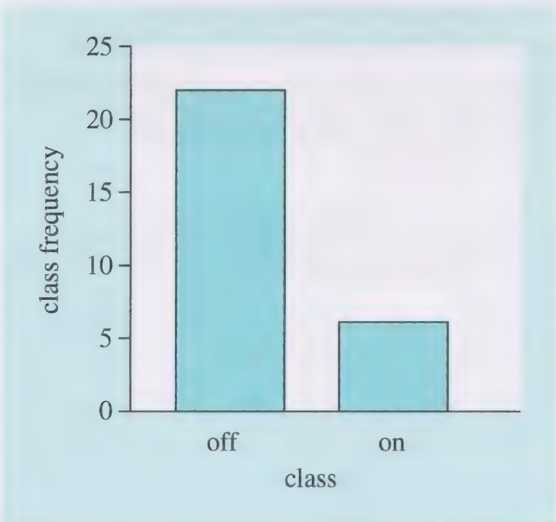


Figure 2.3 Bar chart showing frequencies of computers that were 'off' and 'on'

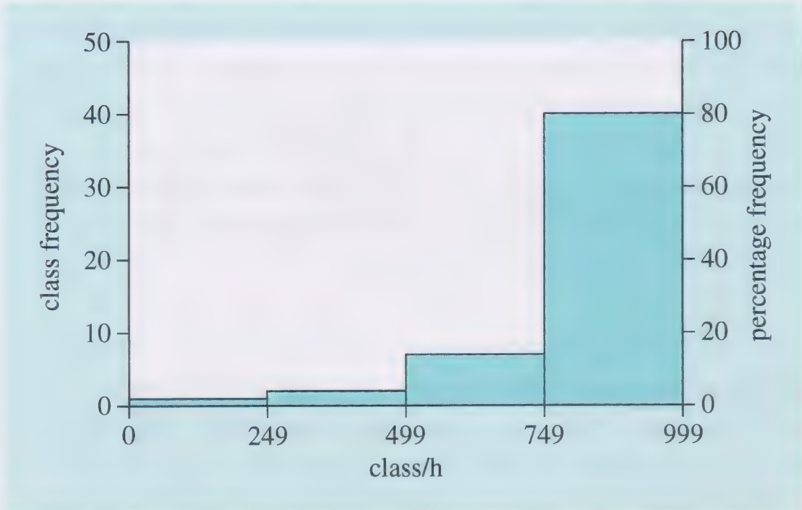


Figure 2.4 Histogram showing absolute and percentage frequencies

A histogram on the other hand is used for quantitative data. Figure 2.4 relates to Activity 2.1. The vertical axis has been used to record the frequency with which the bulbs fall into the different classes. The left-hand vertical axis records the absolute class frequencies, while the right-hand vertical axis records the relative class frequencies as percentages. Using relative (or percentage) frequencies gives a picture that is not affected by the sample size.

Because the classes represent quantitative measurements it is necessary to distinguish between 'class marks' and 'class boundaries'. Class marks denote the range of *measured* values of a given quantity and are recorded with a certain precision. For instance, suppose you measured the lengths of a series of components to the nearest millimetre; you might then have class marks such as 150–154, 155–159, 160–164 mm. However, given the precision of the measurements, the first class might actually contain any length in the range 149.5–154.4 mm, and so on for the other classes. These modified ranges are known as the class boundaries, thereby indicating that with a given set of class marks there exists an implicit set of class boundaries. The difference between two successive class boundaries is known as the class interval; in the example here the class interval is $154.4–149.5 = 4.9$ mm.

Apart from showing an overall picture of a set of data, another benefit of plotting a histogram is that it can help to identify anomalies and errors in the data, such as impossible values.

ACTIVITY 2.2

Table 2.3 shows the frequency distribution resulting from measurements (in centimetres) of the diameters of a series of 200 ball bearings. Draw a histogram showing both the absolute and percentage class frequencies.

Table 2.3 Frequency distribution of ball bearing diameters

Frequency	Class boundaries
2	0.95–0.96
4	0.96–0.97
18	0.97–0.98
44	0.98–0.99
66	0.99–1.00
44	1.00–1.01
16	1.01–1.02
4	1.02–1.03
2	1.03–1.04

2.3 Cumulative frequency diagrams

A histogram shows the frequency of occurrence of each of various classes. However, you might be interested in the cumulative frequency of observations that lie above or below a given value rather than in the frequency of each class. For example, in Activity 2.2 you might have wished to know the total number of ball bearings whose diameter was greater than a desired maximum or less than a desired minimum. Of course, these figures could be calculated from the histogram, but a graphical representation of cumulative frequency is sometimes more useful. Figure 2.5 shows a cumulative frequency chart constructed from the information contained in the histogram in Figure 2.4. The chart contains two curves: (a) is the 'more than' curve and (b) is the 'less than' curve. Each has what may be roughly described as an 'S' shape; this is a characteristic of such curves. (Another name for a cumulative frequency curve is an 'ogive'.) You should note that the two curves give meaningful information only about the number of items that are more than, or less than, each class mark.

2.4 Pareto analysis

Pareto analysis is a technique that is often used in deciding where to focus improvement activities. The craft knowledge that lies behind the technique is the frequently observed phenomenon that in many situations something like 80% of problems are usually attributable to just a 'vital few' sources. Pareto, often referred to as the 80:20 rule, can help you to identify these 'vital few' sources.

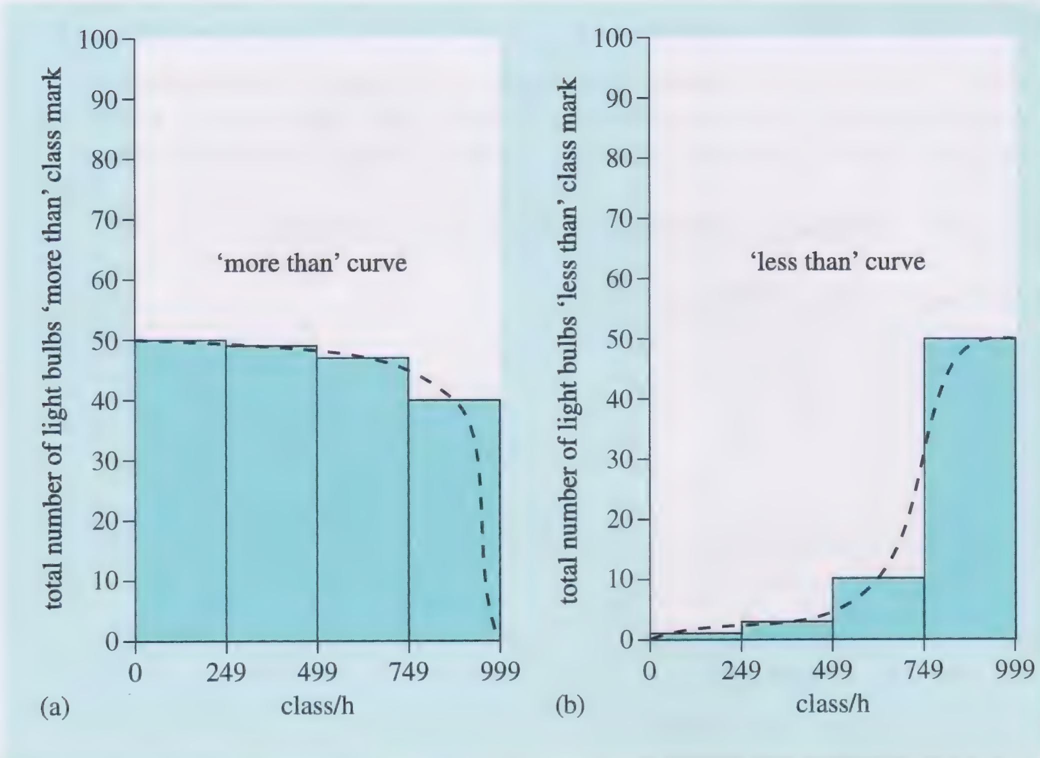


Figure 2.5 Cumulative frequency diagram: (a) 'more than'; (b) 'less than'

Pareto analysis is carried out using a histogram but the classes are ordered in terms of their decreasing frequency. Thus, in a Pareto diagram the extreme left-hand column represents the class with the largest frequency of observations and the extreme right-hand column usually represents the class with the lowest frequency of observations. (I say 'usually' because sometimes the extreme right-hand column denotes a class made up of a large number of small classes that have been grouped together under a heading such as 'other'; and this composite class may be larger than some to its left.) Figure 2.6 depicts a Pareto diagram and also includes a line showing cumulative frequency.

2.5 Scatter diagrams

A scatter diagram is used to look for possible relationships between two variables. In Figure 2.7 the number of customer complaints per 1000 call-outs has been plotted against the call response times. The pattern of the plot indicates that the two variables are associated; a random pattern would have suggested they were not. Because y increases with x the relationship between them is described as direct. Figure 2.8 shows other typical patterns that might be obtained, together with the interpretations that would be placed upon them. Note that establishing a relationship does not necessarily mean there is a causal relationship. (If you are not sure what a causal relationship is, reading Box 2.1 will make it plain.)

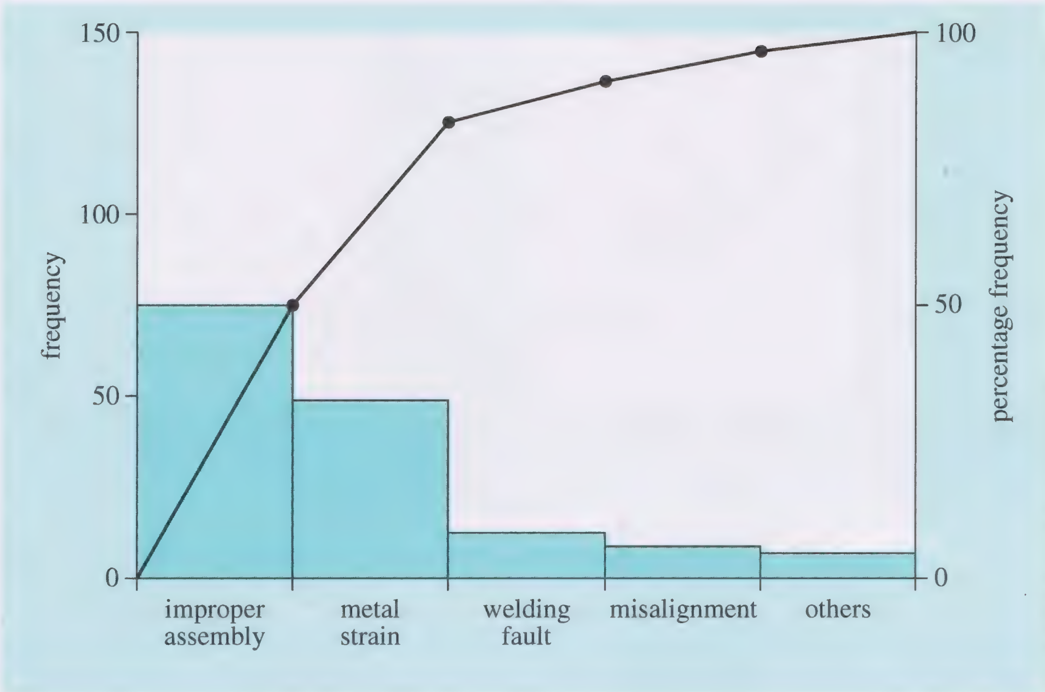


Figure 2.6 Pareto diagram showing absolute and percentage frequencies for different causes of failure and their cumulative frequency

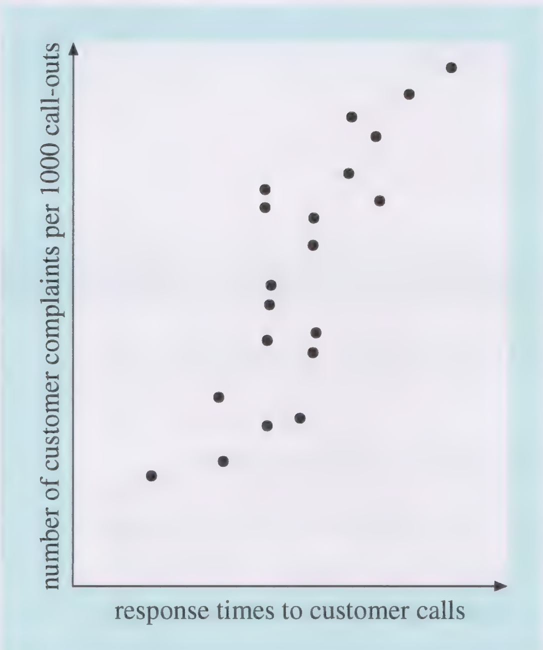


Figure 2.7 Scatter diagram showing a direct relationship between number of customer complaints per 1000 call-outs and call response times

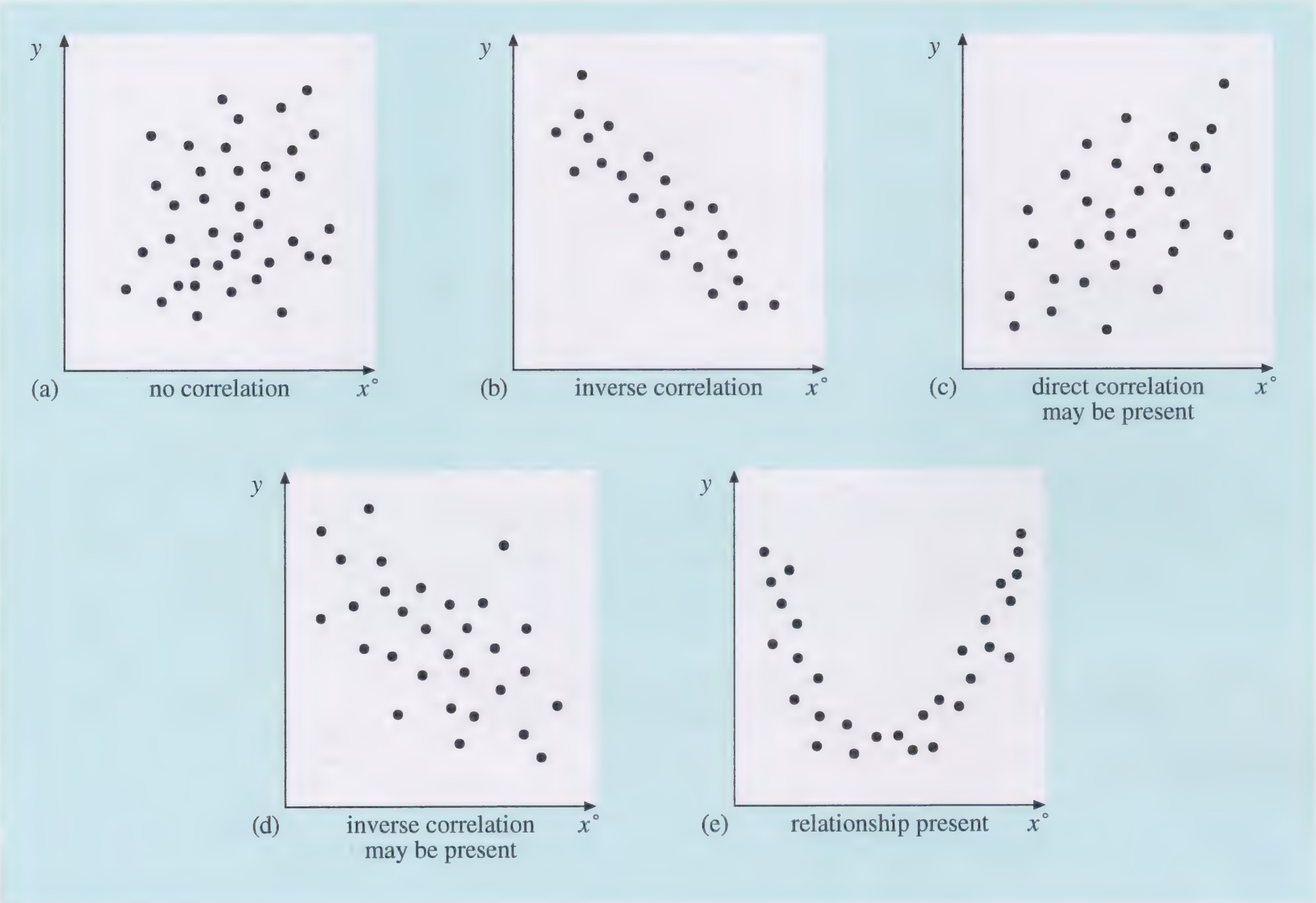


Figure 2.8 Other typical patterns

BOX 2.1 CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with one's shoes on is strongly correlated with waking up with a headache.

Therefore, sleeping with one's shoes on causes headache.

The above example commits the correlation-implies-causation fallacy, as it prematurely concludes that sleeping with one's shoes on causes headache. A more plausible explanation is that *both are caused by a third factor*, in this case alcohol intoxication ...

Source: http://en.wikipedia.org/wiki/Correlation_does_not_imply_causation (accessed 26 January 2007)

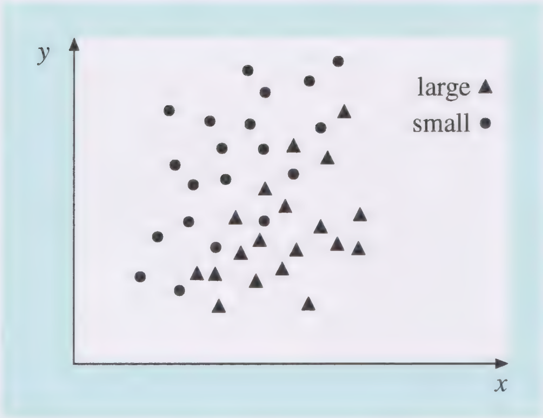


Figure 2.9 Scatter diagram with stratification

Sometimes relationships can be masked because the data plotted are not genuinely homogeneous. For example, Figure 2.8a appears to show no relationship in its present form but, if x represents the amount spent on advertising and y the total number of sales of a product that is available in two sizes, distinguishing between the sales figures for each size might reveal a different picture, as seen in Figure 2.9. Division of the data in this way, so that information from different sources can be looked at separately, is called stratification.

The level of detail at which information should be presented always depends on the intended use. In general, as information becomes aggregated, its usefulness in making small-scale decisions decreases. For example, to decide whether to make a particular process the subject of an improvement exercise, it may be enough to examine the yield figures for the process as a whole. However, in order to tackle the problem and decide which aspects of it to concentrate on, it would be necessary to obtain information about yields at different stages of the production process, and perhaps even yields from individual machines or operating shifts on those machines.

An alternative form of the scatter diagram which can be very useful is a time sequence plot in which a variable is plotted against time. By convention, time is usually shown along the horizontal axis and successive points are connected by straight lines.

ACTIVITY 2.3

The same advertisement was placed on the back page of each issue of a monthly magazine for a year. The number of responses to each advertisement is shown in Table 2.4. Make a time sequence plot of the data.

Table 2.4 Number of responses

Issue	Number of responses
1	120
2	100
3	100
4	110
5	90
6	60
7	80
8	50
9	60
10	30
11	20
12	20

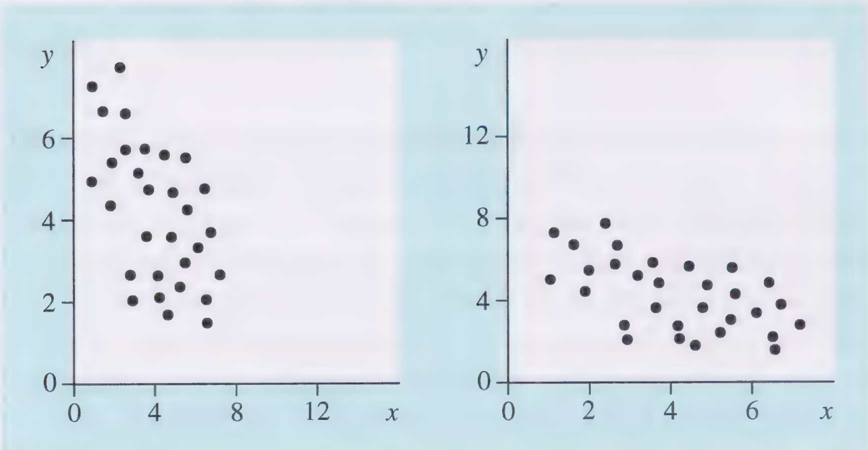


Figure 2.10 Effect of choice of scale

As with many other graphical methods for presenting data, choice of scale is very important when drawing scatter diagrams. Look at Figure 2.10, which shows the same data plotted using different scales. A quick visual examination of each of the two plots would lead to very different conclusions. Part of this problem could be overcome by identifying the maximum and minimum values to be accommodated on each axis before devising the scales, but misinterpretation is still possible if you rely solely on visual impression.

Now do Exercise 2.2 in the Computer Exercises Booklet.

3 SUMMARY MEASURES

The various ways of presenting, classifying and ordering that I described in Section 2 reveal information contained within a set of data. Another way of accessing such information is by using measures that summarise the data numerically. I shall be looking at three of them in this section: location, dispersion and symmetry.

3.1 Measures of location

A measure of location of a set of data is meant to denote a value that is typical or representative of all the data. The most commonly used measures – the mean, median and mode – are all different ways of trying to describe the location of the 'centre' of the data. The similarity in their names is rather confusing but the statistical differences between them are very clear. Mean, sometimes referred to as the arithmetic mean or average, is calculated by summing all the values and dividing the total by the number of values. For example, the mean of the sample 2, 22, 9, 13, 6 is given by:

$$\text{mean} = \frac{2+22+9+13+6}{5} = \frac{52}{5} = 10.4$$

In more formal language the mean is defined as:

$$\text{mean} = \frac{\text{sum of observations}}{\text{number of observations}} \quad \text{or} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where

$\sum\limits_{i=1}^{n}$ (pronounced sigma) means 'the sum of', and the limits $i = 1$ and n indicate that the sum is calculated for each value of x between 1 and n, where n is the total number of observations in the sample

x_i is an observation

$\bar{x}$ (pronounced 'x bar') is the sample mean.

ACTIVITY 2.4

Calculate the means of the following samples:

(a) 2, 6, 4, 9, 5, 7, 7

(b) 3, 8, 2, 9, 12, 36 ●

Whereas the mean is calculated, the median is positioned. When data values are rearranged from smallest to largest, the median is the middle value, or half-way between the two middle values when the number of observations is even. In the example at the beginning of this subsection, where there is an

odd number of observations, the median (the middle value) is 9. To take another example, consider the following observations:

6, 3, 8, 2, 15, 1, 25, 5

Rearranging them in ascending order gives:

1, 2, 3, 5, 6, 8, 15, 25

The number of observations (n) is 8, which is an even number, so the median is calculated from the two middle observations. These are 5 and 6, and the median of the sample is therefore:

$$\text{median} = \frac{5+6}{2} = 5.5$$

ACTIVITY 2.5

What are the medians of the samples given in Activity 2.4? ●

The third measure of location – the mode – is defined as the most frequently observed value in a sample.

For example, the mode of the following sample:

14, 19, 17, 21, 18, 19, 24, 19

is 19. This number occurs three times while the others each appear once. If no observation occurs more than once then the sample has no mode; if two or more observations occur the same number of times (and more frequently than any of the other observations) then the sample has more than one mode and it is said to be multimodal. The following sample:

6, 8, 8, 3, 7, 5, 7, 9

is multimodal as there are two modes, 8 and 7. In the next sample:

1, 8, 5, 7, 13

the frequency of each number is the same and therefore there is no mode.

Unlike the mean and the median, the mode is applicable to both categorical and quantitative data. For example, if you were given a list of paint colours selected by customers, the mode would be the colour selected by most customers. If there were several colours with the highest number then the data would be described as multimodal.

ACTIVITY 2.6

Identify the mode for each of the following samples of observations:

(a) 3, 8, 19, 5, 6, 3, 7

(b) 22, 56, 7, 4, 8, 24, 4, 6. ●

3.2 Measures of dispersion

In this section I shall consider the second type of summary measure, namely, dispersion. The dispersion of a set of data is simply a measure of the spread of the different observations in relation to some given point. Frequently, that chosen point is the mean of the sample, which enables us to say something about the dispersion of observations about the sample mean. A measure of dispersion that is very easy to calculate is known as the sample range, R. It is the difference between the smallest and the largest of a set of observed values and is defined in statistical terminology as:

$$R = x_{\max} - x_{\min}$$

where

R is the sample range

$x_{\max}$ is the maximum observed value of x

$x_{\min}$ is the minimum observed value of x.

Thus if the minimum value is 20 (i.e. $x_{\min} = 20$) and the maximum is 25 (i.e. if $x_{\max} = 25$) then:

$$R = 25 - 20 = 5$$

If the minimum value is –5 and the maximum is 15 then R is 20.

Because the range depends only on two extreme values it can be misleading, especially when comparing the dispersion of two sets of data. Another measure which is often used is the interquartile range (IQR) or mid-range. This is the range within which the middle 50% of the values fall.

Another very widely used measure of dispersion is standard deviation. Essentially, it is calculated by:

1 finding the difference between each value and the mean
2 squaring all these differences
3 adding all the squared differences together and dividing the total by the number of values minus 1
4 and then taking the square root of the result.

The standard deviation is usually written as s and is defined by the following formula:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $(x_i - \bar{x})^2$ is the square of the difference between each value and the mean.

The square of the standard deviation, s^2, is called the variance and can itself be used as a measure of dispersion. The formula for standard deviation can be simplified by writing it as:

$$s = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]}$$

or

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

The calculation of s can then be made easier by breaking it into the following steps:

1 Beginning with a sample x_i where $i = 1, ..., n$, calculate the sum of the squared values of x_i:

$$\sum_{i=1}^{n} x_i^2 = \text{sum of } \left(x_i^2\right)$$

2 Next calculate the mean, $\bar{x}$:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\text{sum of } x_i}{\text{number of observations}}$$

3 Then multiply the sample size, n, by the squared value of the mean: $n\bar{x}^2$.

4 Subtract the total obtained in step 3 from that in step 1 and then divide the result by $(n - 1)$ to give the variance s^2. Finally, find the square root of the variance to obtain the standard deviation.

Now let us look at how these steps work. Consider the following sample:

20, 18, 13, 25, 7, 21, 6, 10

1 Calculate the sum of the squared values of x_i:

x_i	x_i^2
20	400
18	324
13	169
25	625
7	49
21	441
6	36
10	100
	$2144 = \sum x_i^2$

2 Calculate the mean:

$$\text{mean} = \bar{x} = \frac{\sum x_i}{n} = \frac{120}{8} = 15$$

3 Multiply the sample size by the square of the mean:

$$n\bar{x}^2 = 8 \times 15 \times 15 = 1800$$

4 Then subtract the total in step 3 from that in step 1 and divide the result by $(n - 1)$ to obtain the variance:

$$s^2 = \frac{2144 - 1800}{7} = \frac{344}{7} = 49.143$$

and the standard deviation:

$$s = \sqrt{49.1} = 7.010$$

ACTIVITY 2.7

Calculate the variance and standard deviation of the following sample:

10, 50, 21, 14, 5

As you will have just found out from Activity 2.7, calculation of standard deviation is extremely tedious, even for small samples. You will see that there is a big contrast between Activity 2.7 and Exercise 2.3 in the Computer Exercises Booklet even though they are both accomplishing the same task.

Now do Exercise 2.3 in the Computer Exercises Booklet.

3.3 Sample estimates

A lot of statistical analysis uses data that are just a randomly selected subset of all the data that could have been gathered and used. By selecting the data at random from the larger set the hope is that it will be representative of the whole. In statistics terminology, the subset of items is referred to as the sample of the larger set; indeed I have used the term 'sample' throughout Section 3. This larger set is known as the population and is usually denoted by N. The number of items in the sample, denoted n, is known as the sample size. The aim is to make statements about the population on the basis of values for parameters calculated using the sample. That is, estimates of population parameters are made using the values derived from a sample. For example, the mean $\bar{x}$ of a sample may be taken as an estimate of the mean of the population, which is denoted by the Greek letter μ (mu). When the sample mean $\bar{x}$ is used as an estimate for μ it is written as follows:

$$\hat{\mu} = \bar{x}$$

where the ^ sign denotes that the quantity is an estimate. Similarly, $\hat{\sigma}^2 = s^2$ means that the estimate of the population variance σ^2 is given by s^2, the sample variance. (σ is also pronounced sigma; it is the lower-case form of Σ.)

3.4 Measures of symmetry

Symmetry means that data values are similarly distributed on both sides of the median. Asymmetry, which in statistics is called skewness, occurs when values on one side of the distribution are more dispersed than on the other side.

Symmetry is important both practically and statistically. Practically, asymmetry can be both an indicator of problems and a help in diagnosing them. One rule of thumb for problem solving is to investigate the causes of the more dispersed 'tail' of the distribution first, because there the data reveal wider variation. In addition, the more a distribution is skewed, the less valid comparisons between data sets become and the less appropriate it is to use many standard techniques.

Although numerical measures of symmetry can be calculated it is often more practical to assess symmetry from graphs of the data. Skewed data shows up in a histogram when one of the tails either side of the peak is markedly longer than the other. When tails are roughly the same length and the peak is centred, the skewness coefficient is near to zero. If the left-hand tail is longer, skewness is negative and the data are called 'left-skewed'. If the right-hand tail is longer, skewness is positive and the data are called 'right-skewed'.

3.5 Comparing summary statistics

In general, data exploration should always begin by looking at a graphical display of the data, often a histogram. However, histograms can include too much detail, and they are not very useful for comparing two or more samples of data. There are options other than histograms, which focus more upon the useful numerical summaries that you have learned in Section 3. One of the most widely known is the boxplot.

A boxplot is a graphical representation of what is called the 'five-number summary' of a data set. These five numbers are:

1. the minimum value that is not an outlier. An outlier is a value that is more extreme than an adjacent value. Adjacent values are those furthest from the median that are still within a distance of 1.5 times the interquartile range from the end of the box.
2. the lower quartile or first quartile, which is the value that cuts off the lowest 25% of the data
3. the median (sometimes called the second quartile because it is the value that cuts the data in half)
4. the upper quartile or third quartile, which is the value that cuts off the highest 25% of the data)
5. the maximum value that is not an outlier.

It is straightforward to draw boxplots of more than one data set on the same scale, and then to use them to compare important aspects of the distribution of the data sets. Figure 2.11 shows boxplots of two sets of response times. As well as giving a clear picture of the five-number summary, a boxplot also allows you to appreciate any lack of symmetry.

Table 2.5 summarises the features of a boxplot and how boxplots can be used to compare data sets.

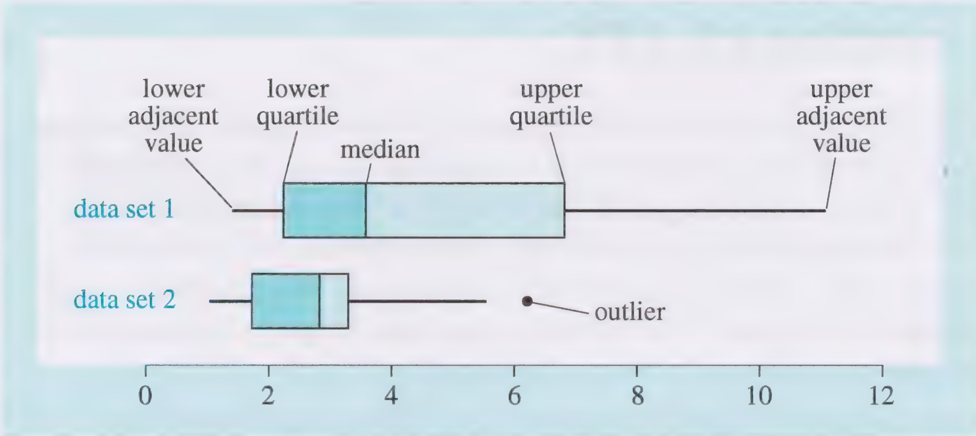


Figure 2.11 Two boxplots

Table 2.5 Features of a boxplot

Measure	Meanings	Comparison	In the example
Location	The vertical line through the box is located at the median.	The relative positions of the medians can be compared.	
Dispersion	The length of the box depicts the interquartile range (IQR).	The lengths of the boxes can be compared.	
	The 'whiskers' either side of the box depict the range. They extend to adjacent values.	The spread across both whiskers and also their individual lengths relative to the box length can be compared. The latter indicates how stretched out the rest of the values are, i.e. the lengths of the tails of distributions.	
	Sometimes values are more extreme than the adjacent values. Often outliers exist.	Potential outliers may need further investigation. Sometimes they are genuine anomalies and sometimes they reveal data-recording errors.	
Symmetry	In asymmetric data the relationships between the upper and lower adjacent values, the upper and lower quartiles, and the median will differ. For example, if the difference between the upper quartile and the maximum is much greater than the difference between the minimum and the lower quartile, then the values of the variable are more skewed to the right.	If the data do not appear to be symmetric, it is possible to determine if each set shows the same kind of asymmetry.	

ACTIVITY 2.8 ....

In the blank column on the right of Table 2.5, add your interpretation of the two data sets in Figure 2.11. ●

4 PROBABILITY

Probability is concerned with the likelihood of events or particular outcomes occurring. Probabilities are expressed as fractions (½, ¼, ¾), as decimals (.5, .25, .75), or as percentages (50%, 25%, 75%). When a probability is expressed as a percentage it can take any value between 0% and 100%. For fractions and decimals a probability can take any value between 0 and 1. A probability of 0 means that something can never happen; a probability of 1 means that something is certain to happen.

In the terminology of probability theory, acts such as coin tossing are known as experiments or trials, and probabilities are written in a particular way that can be a bit offputting. If the letter A refers to a particular event or outcome – for example, obtaining a head in the toss of a coin, obtaining a six in the throw of a die, or your hard disk crashing – the probability that A will occur can be denoted by $P(A)$. The fact that the value of $P(A)$ must be greater than or equal to zero and less than or equal to unity is expressed mathematically as:

$$0 \le P(A) \le 1$$

The classic definition of probability states that if an experiment or trial can result in n equally likely, mutually exclusive outcomes, r of which correspond to the occurrence of event A, then the probability that A will occur is:

$$P(A) = \frac{r}{n} = \frac{\text{no. of events in which } A \text{ occurs}}{\text{total no. of events}}$$

'Equally likely' and 'mutually exclusive' are very important. This formula does not hold without them. Thus, in the toss of a coin, the coin must be fair and the person performing the trial must not cheat. 'Mutually exclusive' means that one, and only one, of the possible outcomes must occur at each trial: tossing a coin results in a head *or* a tail; in throwing a die the outcome is the showing of a single face.

In the tossing of a coin there are only two possible outcomes (therefore $n = 2$) and each event (a head or a tail) is independent and equally likely, so the probability of the outcome being a head (event A, $r = 1$) is:

$$P(A) = \frac{1}{2} = 0.5$$

Similarly, the probability of the outcome being a tail (event B, again $r = 1$) is:

$$P(B) = \frac{1}{2} = 0.5$$

To take another example, consider a pack of 52 well-shuffled playing cards. Each card is unique and any may be cut at random; therefore in this case

$n = 52$. The probability of cutting any specified individual card (event A, $r = 1$) must be:

$$P(A) = \frac{1}{52}$$

The probability of cutting any one of the four aces ($r = 4$) must be:

$$P(A) = \frac{4}{52} = \frac{1}{13} = 0.077$$

The probability of cutting a heart ($r = 13$) is:

$$P(A) = \frac{13}{52} = \frac{1}{4} = 0.25$$

The second important point to note about probability is that the sum of the probabilities of all the possible outcomes must equal 1. (If you think about it, one of the outcomes is certain to occur.) Thus in the case of tossing a coin:

$$P(A) + P(B) = 0.5 + 0.5 = 1$$

ACTIVITY 2.9

(a) What is the probability of throwing a six with a fair die?

(b) What is the probability of cutting a black (spade or club) card from a pack of 52 playing cards?

4.1 The rule of addition

A frequent area of interest is the probability that either one of two independent exclusive events A or B will occur. Independent events are those in which the occurrence of one event does not in any way influence the probability of occurrence of other events. For example, the fact that a coin comes down head in one toss does not influence the result of the next toss. In such cases the probability that either A or B will occur – written as $P(A \text{ or } B)$ – is given by adding together the separate probabilities for each. That is:

$$P(A \text{ or } B) = P(A) + P(B)$$

For example, the probability of a head or a tail is:

$$P(\text{head or tail}) = P(A) + P(B) = 0.5 + 0.5 = 1$$

And the probability of throwing a three or a four with a fair die is:

$$P(3 \text{ or } 4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

4.2 The rule of multiplication

So far I have considered only the probability of occurrence of single events or outcomes but what if I wish to determine the probability of successive events, such as obtaining two successive heads in two tosses of a coin, rolling three sixes, or cutting four aces? Let us consider the coin-tossing example first. If I write down the set of all possible outcomes that can result from a coin being tossed twice I obtain:

head head, tail tail, head tail, tail head

So, the outcome 'head head' (event A) is only one ($r = 1$) of four ($n = 4$) possibilities. Therefore:

$$P(A) = \frac{1}{4}$$

This is an easy example, but if I draw up a list of the possible combinations of outcomes resulting from the rolling of a die twice (or a pair of dice once) the exercise starts to become rather tedious. The following list shows the possible combinations:

1 1	2 1	3 1	4 1	5 1	6 1
1 2	2 2	3 2	4 2	5 2	6 2
1 3	2 3	3 3	4 3	5 3	6 3
1 4	2 4	3 4	4 4	5 4	6 4
1 5	2 5	3 5	4 5	5 5	6 5
1 6	2 6	3 6	4 6	5 6	6 6

There are 36 possible outcomes ($n = 36$), in which the desired or favourable outcome of two sixes occurs only once ($r = 1$). Thus, the probability of throwing two successive sixes is 1/36.

Obviously, if I wanted to calculate the probability of rolling three (or more) successive sixes the task would quickly get out of hand: a quicker way of calculating such probabilities is needed. Fortunately, this is provided by a rule that is known as the multiplication law of probability. It is important to note that this rule applies only to independent events, that is, situations in which the occurrence of one event has no influence on the occurrence of any following events. For example, the outcome of the first toss of a coin does not in any way influence the outcome of the second toss. Similarly, once a card has been cut from a pack it is replaced, and so this event has no influence on subsequent events. According to the multiplication law, if A denotes the first of two independent events, and B the second, then the probability of A followed by B, i.e. $P(AB)$, is the probability of the first, $P(A)$, multiplied by the probability of the second, $P(B)$:

$$P(AB) = P(A) \cdot P(B)$$

(*Note:* the '·' in the formula stands for multiply.)

Checking this formula with the example of rolling two successive sixes with a die gives:

$$P(66) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

By extension, the probability of throwing three successive sixes would be:

$$P(666) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$$

ACTIVITY 2.10 ....

Calculate the probability of:

(a) cutting two successive aces if the first card is returned to the pack before the second one is selected.

(b) cutting one ace and then any heart if the first card is returned to the pack before the second one is selected.

4.3 Conditional probability

When events are not independent another formula must be used for calculating the probability of multiple events. To illustrate this type of problem I shall return to the example of a pack of playing cards. Suppose that, instead of cutting the pack, the trials consisted of drawing a single card out of the pack and not replacing it. Following this procedure, what would be the probability of drawing two successive aces? The following formula applies:

$$P(AB) = P(A) \cdot P(B \mid A)$$

In this case, the probability of A followed by B, $P(AB)$, is the probability of A multiplied by the probability of 'B given A', written $P(B|A)$. It is necessary to use $P(B|A)$ rather than $P(B)$ because the occurrence of A influences the probability of occurrence of B. In other words, if you have taken one ace from a pack and not put it back this changes the probability of drawing an ace on your second draw. In statistical terminology this is referred to as 'the probability of B conditional upon A'.

For the first draw $n = 52$ and $r = 4$ (there are four aces in the pack), so the probability of drawing one ace is 4/52 or 1/13. If an ace was selected in the first draw then in the second draw, because the number of cards is reduced to 51 ($n = 51$) and only three aces are left in the pack ($r = 3$), the probability of drawing a second ace must be:

$$P(B \mid A) = \frac{3}{51} = \frac{1}{17}$$

The probability of drawing two successive aces is thus:

$$P(AB)=\frac{1}{13}\cdot\frac{1}{17}=\frac{1}{221}$$

ACTIVITY 2.11

Calculate the probability of drawing two successive clubs if the first card is not returned to the pack before the second one is selected.

The final case I shall look at is the calculation of the probability that event A or event B or both A and B simultaneously will occur. In such cases A and B are not mutually exclusive. This form of probability is written as $P(A+B)$. For example, consider the probability of drawing a heart (event A) or a court card – jack, queen, or king – (event B) or both, that is, a heart that is a court card. In these cases there are three quantities to consider:

- the probability of drawing a heart without considering the court cards at all
- the probability of drawing a court card without considering the hearts
- and the probability of drawing a court card that is a heart.

These are combined using the formula:

$$P(A+B)=P(A)+P(B)-P(AB)$$

Given the formula for cases of conditional probability, this can be written as:

$$\begin{aligned}P(A+B)&=P(A)+P(B)-P(A)\cdot P(B|A)\\&=\frac{13}{52}+\frac{12}{52}-\left(\frac{13}{52}\cdot\frac{3}{13}\right)\\&=\frac{11}{26}\end{aligned}$$

Because it is easy to visualise a pack of cards an easy way to check this result is to count up the number of hearts in a pack and the number of court cards that are not hearts. This gives 22 ways in which a heart or a court card can occur. Therefore, the probability of occurrence is $22/52=11/26$, as before.

ACTIVITY 2.12

What is the total probability of obtaining an even number or a number greater than four, or both, when throwing a die?

5 PROBABILITY DISTRIBUTIONS

A probability distribution is a curve that shows all the values that a variable can take and the likelihood that each will occur. In this section I shall be looking at three particular types of probability distribution: the normal, from which the standard normal distribution derives; the binomial; the Poisson. Each of these has a mathematical formula that relates the values of a variable (or quantity) with the probability of observing those values in the population of the variable. I shall start by looking at the normal distribution.

5.1 The normal distribution

The frequency distributions or histograms of many continuous variable quantities are bell-shaped and approximate to what is known as the normal distribution. (I say 'approximate' because the normal curve is exactly symmetrical about the mean value of the population and its two tails get closer and closer to the bottom, or horizontal, axis but never actually touch it.) This is illustrated in Figure 2.12. For example, if you took a large sample of light bulbs and measured their respective times to failure, classified these and drew up a histogram, it would look roughly like the normal curve. The larger the sample, the closer would be the fit.

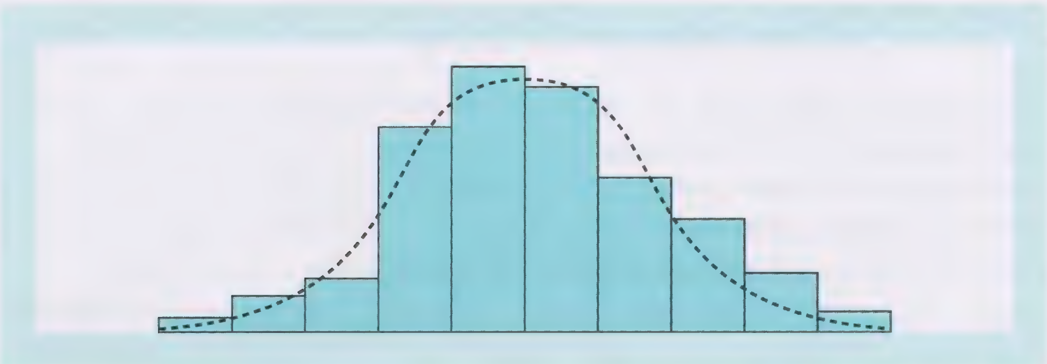


Figure 2.12 Histogram approximated by the bell-shaped normal curve

There are some other interesting observations to be made about normally distributed data. First, the mean of a normal population will occur at the peak of the curve; moreover, the mean, median and mode will coincide. Second, there is a very important relationship between the standard deviation of a population and the normal curve. When the standard deviation is calculated for a normal frequency distribution, 68.3% of all the readings in the distribution will occur between plus and minus one standard deviation of the mean ($\mu \pm 1\sigma$), 95.4% will occur between plus and minus two standard deviations of the mean ($\mu \pm 2\sigma$), and 99.7% between plus and minus three standard deviations of the mean ($\mu \pm 3\sigma$), as shown in Figure 2.13. Thus,

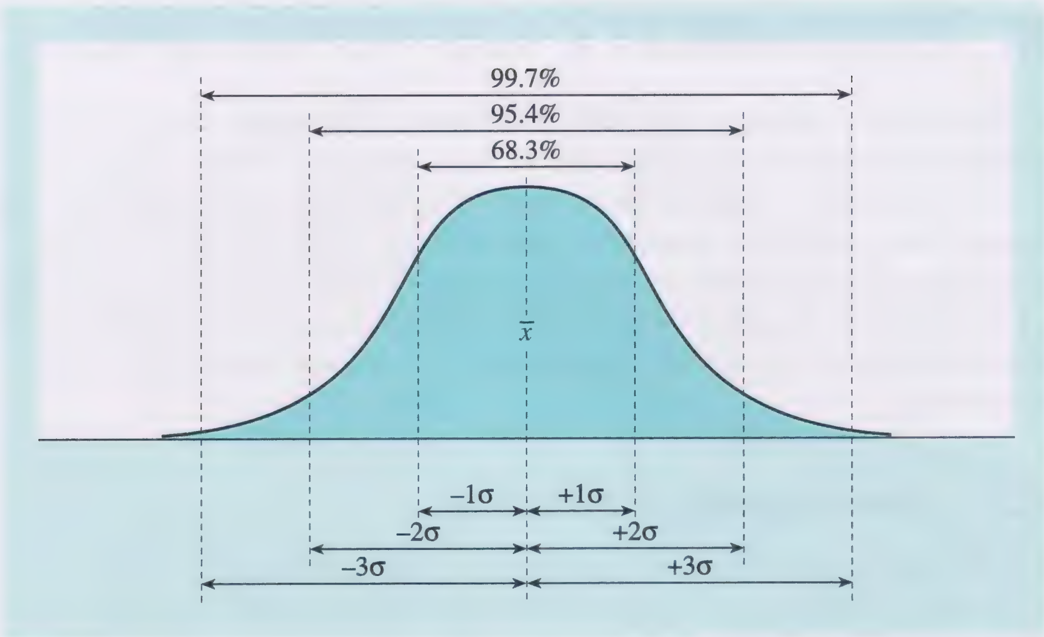


Figure 2.13 Percentages of the normal distribution

if you know the mean and standard deviation for a normal distribution, it is possible to calculate the following:

- the percentage of values that will fall between any two readings of different values
- the total amount of variation that may, for all practical purposes, be expected from that distribution, i.e. $\mu \pm 3\sigma$.

The major justification for the use of the normal distribution as an approximation to many real situations stems from an important theorem which is known as the central limit theorem. This theorem states that for a given population, irrespective of the shape of its probability distribution (it might look extremely non-normal), the distribution of mean values $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_n)$ of samples of size n drawn from the population will tend to a normal distribution as the size of n increases. However, having gathered a data set, it is wise to check whether it approximates to a normal distribution rather than just assume it does or does not. I shall look at two ways of doing this here.

The first way is to check for normality by using the data to plot a histogram. The second is to construct a normal probability plot. This is a graphical technique where the data are plotted (on the vertical axis) against a theoretical normal distribution (on the horizontal axis); if the resulting points lie roughly along a straight line, then a normal model is plausible. The methods that can be used for the manual derivation of the values needed for the horizontal axis are beyond the scope of this course but, as Exercise 2.4 in the Computer Exercises Booklet shows, a normal probability

plot can easily be produced using the statistical software supplied with the course.

Now do Exercise 2.4 in the Computer Exercises Booklet.

5.2 The standard normal distribution

While different normal variables will have distribution functions with a similar shape, the exact height and width of each distribution will depend on the population mean and the standard deviation. However, there is a special form of the normal curve which is chosen so that its mean, μ, is equal to 0 (zero) and its standard deviation, σ, is equal to 1. This is known as the standard normal distribution. The total area under the standard normal curve is the sum of the probabilities for each value in the population and must therefore be equal to 1. Further, the area under the curve between any two given values represents the probability of making a random observation in that range. Areas under the standard normal curve are given in tables, which appear in one of two forms:

1 which gives the areas between the mean and given values of z as shown in Figure 2.14
2 which gives the tail areas, i.e. the areas beyond the given values of z as shown in Figure 2.15.

The areas shown in Figures 2.14 and 2.15 must add up to 0.5 for a standard normal curve because they account for half of the total probability of 1. Therefore, subtracting the tail area from 0.5 gives the area between the mean and z, and vice versa. Because any normal variable can be standardised, these tables of areas can be used to calculate the probabilities of observing any particular values for items in a population.

A normal random variable X with mean μ and standard deviation σ can be standardised as follows:

$$z = \frac{X - \mu}{\sigma}$$

where z is known as the standard normal variable of X.

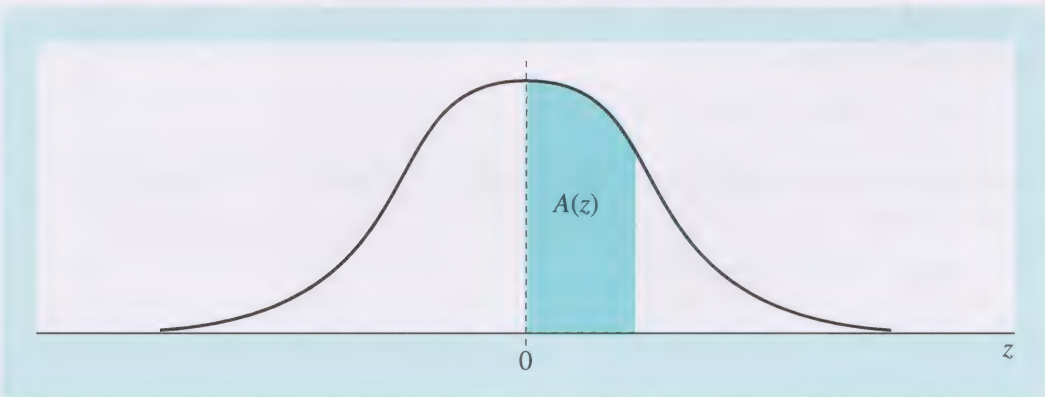


Figure 2.14 Area under the standard normal curve

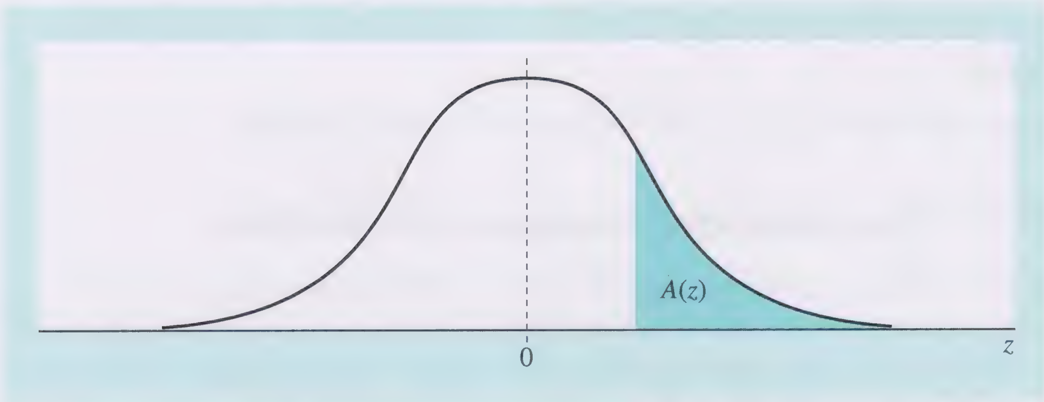


Figure 2.15 Area in the tail of a standard normal curve

A table for areas in the tail of a standard normal curve is given in the Appendix. Suppose you are interested in some quantity (for example lifespan or length) concerning items in a given population. If you choose a value for z, say z_1, and look up the value $A(z)$ alongside it, then you have the area under the standard normal curve in the tail area beyond z_1 or, in other words, the probability that an item drawn at random from a population will have a value beyond z_1. For example, if $z = 1.93$ then, from the Appendix, $A(1.93) = 0.0268$. So the probability that an item chosen at random from the population will have a value beyond z_1 is 0.0268.

To take a practical example, suppose you have a batch of 2000 light bulbs with an average burning life of 1000 hours and a standard deviation of 200 hours. It would be interesting to ask how many light bulbs might be expected to fail, say, before 700 burning hours. The first thing to do is to calculate the value of z in the expression $z = X - \mu / \sigma$ in order to standardise the data. In this case the value of X is 700, μ is 1000, and σ is 200, which gives:

$$z = \frac{700 - 1000}{200} = -1.5$$

Ignore the negative sign and let $z = 1.5$.

Next, find the area under the standard normal curve beyond $z = 1.5$. This area corresponds to the probability that a light bulb will have a burning life of less than 700 hours, that is:

$$P(X < 700) = A(1.5)$$

Looking up the standard normal table with $z = 1.5$ gives $A(1.5)$ equal to 0.0668. Thus:

$$P(X < 700) = 0.0668$$

Finally, since this is the probability that a single, randomly chosen, light bulb will burn for less than 700 hours, the total number that might be expected to fail before this time must be:

$$0.0668 \times 2000 = 134$$

ACTIVITY 2.13

Delivery times from receipt of order are normally distributed with a mean of 100 days and a standard deviation of 15 days. What is the probability that:

(a) a delivery will occur in less than 90 days?

(b) a delivery will take longer than 130 days?

5.3 The binomial distribution

The normal distribution is a continuous distribution but for many histograms the heights of the columns are calculated by simply counting the frequencies of various classes or categories, and thus give rise to discrete distributions.

In investigating some problems there are only two categories of interest – for example, conforming and non-conforming or meeting a target and not meeting a target – where the number of items in each of the two categories is a whole number. In these cases it is possible to draw on a probability distribution known as the binomial distribution. This distribution can accurately represent the probable frequency of finding a specified number in one and/or the other category in a sample of a given size if you know the probability for the population as a whole.

Consider a large set of data relating to the arrival times of trains where 20% of the trains are judged not to have arrived 'on time'. If you were to look at just one train at random, then the probability, p, that it would not have arrived 'on time' (event A) would equal 0.2. Conversely, the probability, q, that it would have arrived 'on time' (event B) would equal 0.8. Since the train would be either not 'on time' or 'on time':

$$p + q = 1$$

What if you were to look at two trains from the same set of data? (It is important in this context that the set of data is so large that the selection of a not 'on time' train does not, for practical purposes, change the proportion of not 'on time' trains in the set of data.) From your knowledge of the laws of probability you can say that the chance that both would not have arrived 'on time' would be:

$$P(AA) = p^2 = 0.04$$

and the probability that both items would have arrived 'on time' would be:

$$P(BB) = q^2 = 0.64$$

Next, what of the probability of picking one train that does not arrive 'on time' followed by one train that does arrive 'on time', or vice versa? Again, using the multiplication law:

$$P(AB) = p \cdot q = 0.16$$

$$P(BA) = q \cdot p = 0.16$$

Thus, the total probability of picking one that arrives 'on time' and one that does not (in any order) would be $2pq$ or 0.32. In tabular form:

Result:	Both not 'on time'	One 'on time' and one not 'on time'	Both 'on time'
Probability:	p^2	$2pq$	q^2

If you are familiar with the algebra topic of binomial expansion you will recognise that these terms are the expansion of $(p+q)^2$. (If you are not familiar with this topic you may find the explanation in Box 2.2 helpful.)

BOX 2.2 BINOMIAL EXPANSION OF $(p+q)^2$

To expand $(p+q)^2$ first write it as:

$$(p+q)^2 = (p+q)(p+q)$$

The expansion is then carried out in the following way. Taking the first term in the first bracket (on the right-hand side of the = sign), multiply it with each of the terms in the second bracket to obtain:

$$p^2 + pq$$

The next step is to take the second term from the first bracket and do the same, obtaining:

$$qp + q^2$$

Finally, adding the two sets of terms together and ordering them, beginning with the highest power of p on the left and ending with the highest power of q on the right, gives:

$$(p+q)^2 = p^2 + 2pq + q^2$$

If you selected three trains from the set of data, the terms would be given by the expansion of $(p+q)^3$, for four trains $(p+q)^4$, and so on.

In general, the binomial distribution of a sample is given by the expansion of $(p+q)^n$, where n is the number of items in the sample drawn at random from a population whose proportion of items in one category is p and whose

proportion of items in the other category is q. The probabilities of drawing 0, 1, 2, 3, 4, etc. items from the first category are given by the successive terms (from right to left, i.e. from the highest power of q to the highest power of p) in the expansion.

To illustrate the use of the binomial distribution, consider the following example. Suppose three items are drawn from a population in which the probability that they do not conform to your standard is p, where $p = 0.1$, and the probability that they do conform to your standard is q, where $q = 0.9$. In this case the value of n in the expansion $(p+q)^n$ is 3. Hence the first task is to expand $(p+q)^3$:

$$\begin{aligned}(p+q)^3 &= (p+q)^2(p+q)\\ &= (p^2+2pq+q^2)(p+q)\\ &= p^3+3p^2q+3pq^2+q^3\end{aligned}$$

The first term (extreme left, highest power of p) gives the probability of drawing n (i.e. 3) non-conforming items. The second term gives the probability of drawing $(n-1)$ non-conforming items, and so on.

Thus:

probability of drawing 3 non-conforming items:	$p^3 = (0.1)^3$	=	0.001
probability of drawing 2 non-conforming items:	$3p^2q = 3(0.1)^2\ (0.9)$	=	0.027
probability of drawing 1 non-conforming item:	$3pq^2 = 3(0.1)\ (0.9)^2$	=	0.243
probability of drawing 0 non-conforming items:	$q^3 = (0.9)^3$	=	0.729
Total		=	1.000

(*Note:* the total probability must equal 1 because all possible outcomes have been covered.)

ACTIVITY 2.14 ....

Calculate the probabilities of drawing 4, 3, 2, 1, and 0 non-conforming items in a sample of four items drawn at random from a population whose proportion of non-conforming items $p = 0.1$, and whose proportion of conforming items $q = 0.9$. Note:

$$(p+q)^4 = p^4+4p^3q+6p^2q^2+4pq^3+q^4$$

Another general formula for the binomial distribution is given by:

$$P(X)=\frac{n!}{x!(n-x)!}p^x\,(1-p)^{(n-x)}$$

where, using the same example of non-conforming and conforming items:

$P(X)$	is the probability of drawing x non-conforming items
n	is the sample size
p	is the proportion of non-conforming items in the population
$(1-p)$	is the proportion of conforming items in the population (q)
!	stands for factorial; for example 2! is factorial 2.

Factorials are enumerated as follows:

$0!=1$

$1!=1$

$2!=2\times 1=2$

$3!=3\times 2\times 1=6$

$4!=4\times 3\times 2\times 1=24$ and so on.

For example, if $p=0.9$ and $q=0.1$, the probability that three items ($x=3$) in a sample of three ($n=3$) are all non-conforming is given by:

$$P(3)=\frac{3!}{3!(3-3)!}p^3\,(1-0.1)^{(3-3)}$$

$$=\frac{3!}{3!0!}p^3\,(0.9)^0$$

$$=p^3=(0.1)^3=0.001 \text{ (as obtained before)}$$

(*Note:* if any base is raised to the power 0 (zero) the answer is 1, e.g. $1^0=1$, $25^0=1$.)

ACTIVITY 2.15 ....

Using the general formula for the binomial distribution, calculate the probability of obtaining five non-conforming items in a sample of five items drawn at random from a population. As before, assume that $p=0.1$ and $q=0.9$. What is the probability of obtaining four non-conforming items? ●

5.4 The Poisson distribution

When describing the binomial distribution I was concerned with a sample of specified size (n) taken from a large population where each item in the sample fitted into one of two categories. There are situations, however, where it may be possible to count the number of times an event occurred but have no means of knowing how many times the event did not occur. To take an example, in examining a length of railway line you could count the number

of times you found a problem, but you could not meaningfully say how many times you did not find one. Similarly, in searching for blemishes on the painted surface of a car door, you could say how many times blemishes were detected, but not how many times they were not detected. In such cases you are dealing with events in a continuum: with the railway line the continuum is one of length, with blemishes it is one of area. The binomial distribution cannot be used in cases like these because the value of n in the fundamental expression $(p+q)^n$ is unknown. However, a probability distribution that can be used, provided that the events of interest are relatively rare, is the Poisson.

If you know the total number of events in a population you can calculate the average number of events you can expect in a sample drawn from that population. The expression for the Poisson distribution allows you to use this value for the average, denoted by m, to calculate the probability of observing the occurrence of 0, 1, 2, 3, 4, etc. events. These probabilities are given by the successive terms of the expression:

$$e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\frac{m^4}{4!}+\ldots\right)$$

where e is a mathematical constant ($e=2.7183$ to four decimal places).

If you had the results of a series of railway line inspections and knew the average number of problems encountered, the expression would tell you how many times you would expect to find different numbers of problems in further inspections. There is only one condition, and that is that m should not vary from trial to trial. For example, in examining a length of railway line for problems the same length of line should always be used.

One of the classic examples that is often used to illustrate the usefulness of the Poisson distribution comes from nineteenth-century records concerning the number of German cavalrymen killed by horse-kicks. During the course of some twenty years, data from ten army units yielded the figures shown in Table 2.6.

Table 2.6 Frequency distribution of deaths of German cavalrymen

Deaths	Number of times this number of deaths occurred
0	109
1	65
2	22
3	3
4	1
5	0
6	0

The total number of deaths was 122, and the total number of readings was 200 (that is, 20 years times 10 army units). Therefore the value of m, in this case the average number of deaths per unit per year, was:

$$m = \frac{122}{200} = 0.61$$

The value of $e^{-0.61}$ is approximately 0.5434; hence, according to the Poisson distribution, the probabilities of finding 0, 1, 2, 3, etc. deaths per year are given by the successive terms in the expansion:

$$0.5434\left(1 + 0.61 + \frac{(0.61)^2}{2!} + \frac{(0.61)^3}{3!} + \frac{(0.61)^4}{4!} + \frac{(0.61)^5}{5!} + \frac{(0.61)^6}{6!} + \ldots\right)$$

Table 2.7 shows the results. As you can see, the fit between the calculated and observed frequencies is remarkably good.

Table 2.7 Comparison of predicted results using the Poisson distribution and actual results in the German cavalry example

Number of deaths	**Probability**	**Poisson frequency expected in 200 readings (to 2 decimal places)**	**Actual frequency**
0	0.5434	108.68	109
1	0.3315	66.30	65
2	0.1011	20.22	22
3	0.0206	4.11	3
4	0.0031	0.63	1
5	3.8×10^{-4}	0.08	0

A general formula for the Poisson distribution is given by:

$$P(X) = \frac{m^x e^{-m}}{x!}$$

where

$P(X)$ is the probability of observing x events

m is the mean or average number of events.

Going back to the cavalry example, the probability of observing three deaths is:

$$P(3) = \frac{(0.61)^3 e^{-0.61}}{3!}$$

$$= \frac{(0.61)^3 \times 0.5434}{3!}$$

$$= 0.0206$$

which is the same value as calculated before.

In certain circumstances, the Poisson distribution is a good approximation to the binomial distribution. For example, in a test in which the probability of finding a non-conforming item in a sample is rather small (for example, less than 1 chance in 10) but constant, then it is considered to be a relatively rare event and suitable for treatment using the Poisson distribution. This approximation is also valid in cases where the number of items or observations is large (greater than 16), or when the sample is small compared with the size of the population (less than 10%).

ACTIVITY 2.16

In looking at the accuracy of 50 sets of invoices, errors were found with the frequencies shown in Table 2.8.

Table 2.8

Number of errors	Frequency
0	0
1	0
2	0
3	1
4	4
5	3
6	6
7	7
8	7
9	5
10	2
11	5
12	2
13	3
14	1
15	2
16	1
17	0
18	1

Using these data, calculate:

(a) the total number of errors

(b) the average or expected number of errors per set of invoices

(c) the probability (according to the Poisson formula) of observing each number of errors, and the Poisson frequency of each number of errors (up to 11) expected in 50 sets of invoices. ●

Now do Exercise 2.5 in the Computer Exercises Booklet.

6 CONCLUSION

This block has looked at different ways in which data can be presented so that their meaning can be understood more easily and quickly, and has presented some basic statistics. It has also introduced you to the statistical software package that you will be using again in the next block. You will find that Block 3 'Techniques' follows on very closely from this one because it starts with the seven 'old' tools of quality improvement, and five out of those seven are techniques you have already practised here.

GLOSSARY

n	sample size
N	size of a population
P	probability
R	range
s	standard deviation of a sample
s^2	variance of a sample
x_i	observed value of some variable
$\overline{x}$	(x bar); mean; the average of a set of numbers
X	a random variable
z	a standardised variable; $z = (x - \mu)/\sigma$
μ	mu; mean of a population
σ	sigma; standard deviation of a population
σ^2	sigma squared; variance of a population
Σ	capital sigma; sum of
$\hat{}$	used above another symbol to denote an estimated value; e.g. $\hat{s}^2$ is an estimate of the sample variance
$\sqrt{}$	square root

ANSWERS TO ACTIVITIES

Activity 2.1

class/h													class frequency
0–249	|												1
250–499	||												2
500–749	𝍸	||											7
750–999	𝍸	𝍸	𝍸	𝍸	𝍸	𝍸	𝍸	𝍸					40

Figure 2.16 Tally chart for light bulb sample

Activity 2.2

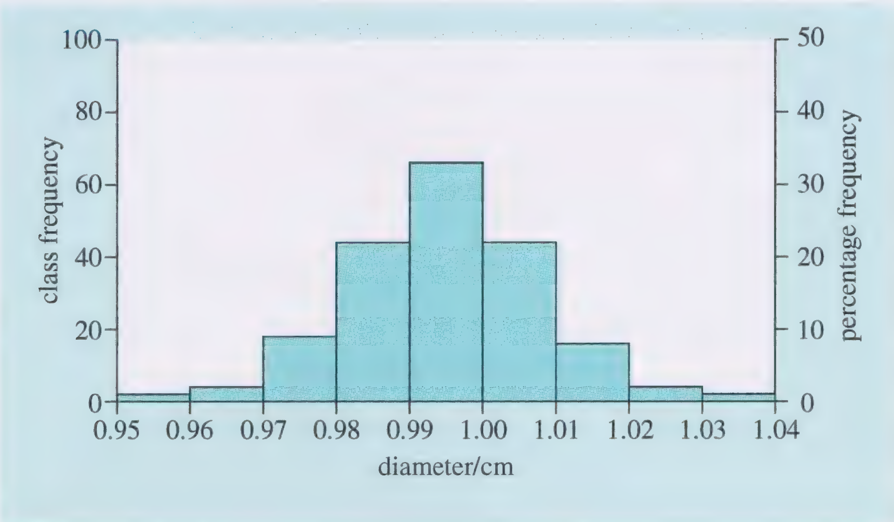


Figure 2.17 Histogram for ball bearing diameters

Activity 2.3

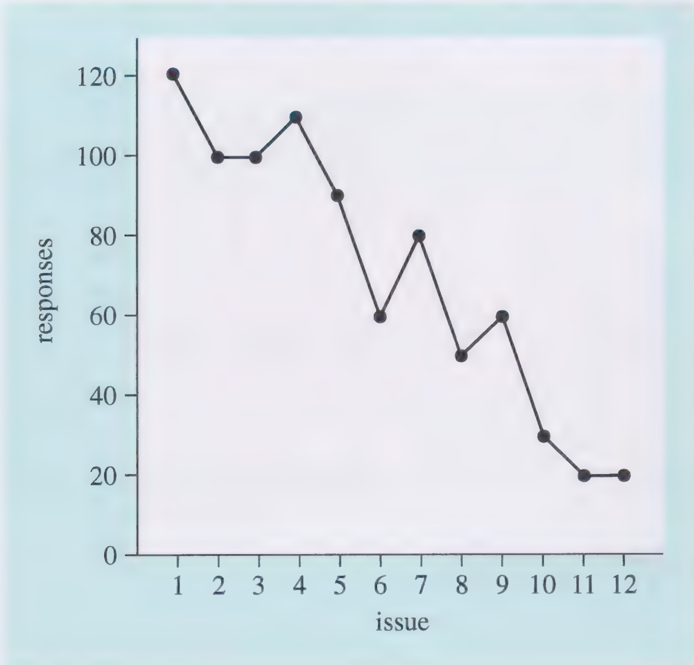


Figure 2.18 Time sequence plot of the data in Table 2.4

Activity 2.4

(a) $\dfrac{40}{7}$ or 5.714

(b) $\dfrac{70}{6}$ or 11.667

Activity 2.5

(a) 6

(b) $\dfrac{17}{2}$ or 8.5

Activity 2.6

(a) 3

(b) 4

Activity 2.7

$$\sum x_i^2 = 3262$$
$$\bar{x} = 20$$
$$n\bar{x}^2 = 2000$$
$$\text{Therefore } s^2 = \frac{3262 - 2000}{4}$$
$$= 315.5$$
$$s = 17.762$$

Activity 2.8

My interpretations are as follows.

In the example:

Location:

- The medians are not far apart.

Dispersion:

- Data set 1 is more dispersed. Its histogram would be flatter.
- Data set 1 has a much larger range.
- One unusually high value appears in data set 2. It is not very far from the rest, and is less than several values in the other data set, so it is likely to be correct, shown up only by the tight grouping of its companions.

Symmetry:

- Both data sets are asymmetric, right-skewed. In fact their skewness coefficients are identical. Despite this, they clearly have different distributions because their medians occupy different positions within both the interquartile and overall ranges.

Activity 2.9

(a) $\frac{1}{6}$

(b) $\frac{26}{52} = \frac{1}{2}$

Activity 2.10

(a) $\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$

(b) $\frac{4}{52} \cdot \frac{13}{52} = \frac{1}{52}$

Activity 2.11

$$\frac{13}{52} \cdot \frac{12}{51} = \frac{3}{51} \text{ or } 0.0588$$

Activity 2.12

$$\frac{3}{6}+\frac{2}{6}-\left(\frac{3}{6}.\frac{2}{6}\right)=\frac{5}{6}-\frac{1}{6}=\frac{4}{6}=\frac{2}{3}$$

Activity 2.13

(a) X is 90, μ is 100 and σ is 15

$$z=\frac{90-100}{15}=-0.67$$

$P(X<90)=0.2514$

(b) $$z=\frac{130-100}{15}=2$$

$P(X>130)=0.02275$

Activity 2.14

Probability of 4 non-conforming items is:	p^4	=	$(0.1)^4$	=	0.0001
Probability of 3 non-conforming items is:	$4p^3q$	=	$4(0.1)^3\,(0.9)$	=	0.0036
Probability of 2 non-conforming items is:	$6p^2q^2$	=	$6(0.1)^2\,(0.9)^2$	=	0.0486
Probability of 1 non-conforming item is:	$4pq^3$	=	$4(0.1)\,(0.9)^3$	=	0.2916
Probability of 0 non-conforming items is:	q^4	=	$(0.9)^4$	=	0.6561
Total				=	1.0000

Activity 2.15

$$P(5)=\frac{5!}{5!(5-5)!}(0.1)^5\,(0.9)^{5-5}=(0.1)^5=0.00001$$

$$P(4)=\frac{5!}{4!(5-4)!}(0.1)^4\,(0.9)^{5-4}=\frac{5!}{4!}(0.1)^4\,(0.9)=0.00045$$

Activity 2.16

(a) The total number of errors is 436.

(b) The average or expected frequency is $\frac{436}{50}=8.72$

(c) $$P(X)=\frac{(8.72)^x\,\mathrm{e}^{-8.72}}{x!}$$

So for two errors, for example:

$$P(2)=\frac{(8.72)^2\,\mathrm{e}^{-8.72}}{2!}$$

$$=\frac{(8.72)^2\times 0.000164}{2}$$

$$=0.0062$$

Table 2.9

Number of errors	Probability	Poisson frequency in 50
0	0.0002	0.01
1	0.0014	0.07
2	0.0062	0.31
3	0.0181	0.90
4	0.0393	1.97
5	0.0686	3.43
6	0.0997	4.99
7	0.1242	6.21
8	0.1354	6.77
9	0.1312	6.56
10	0.1144	5.72
11	0.0902	4.51

APPENDIX: AREAS IN TAIL OF THE NORMAL DISTRIBUTION

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
3.0	0.00135									
3.1	0.00097									
3.2	0.00069									
3.3	0.00048									
3.4	0.00034									
3.5	0.00023									
3.6	0.00016									
3.7	0.00011									
3.8	0.00007									
3.9	0.00005									
4.0	0.00003									

Reproduced with permission from Murdoch, J. and Barnes, J. A. (1985) *Statistical Tables for Science, Engineering, Management and Business Studies* (2nd edn), Macmillan Publishers Ltd.

ACKNOWLEDGEMENTS

Grateful acknowledgement is made to the following sources:

Appendix

Murdoch, J. and Barnes, J.A. (1985) *Statistical Tables for Science, Engineering Management and Business Studies*, Macmillan Publishers Ltd.

T889
Problem solving and improvement: quality and other approaches

Study guide

Block 1 Introduction

Block 2 Statistics

Block 3 Techniques

Block 4 Methods and approaches

Block 5 Managing problem solving and improvement

Index of techniques

ISBN 978 0 7492 2349 6

Cover image from www.stockxpert.com